Volume 64 Number 3 2026

NATIONAL SOCIAL SCIENCE JOURNAL

Official Journal of the National Social Science Association

NATIONAL SOCIAL SCIENCE ASSOCIATION

EST. 1975

Name of Publication: **NATIONAL SOCIAL SCIENCE JOURNAL**
Issue: Volume 64 Number 3, 2026

Office of Publication: Austin Peay State University
College of Education
Clarksville, TN 37040

Online journals: https://nssascholars.org/publications/

The National Social Science Technology Journal is abstracted in Cabell's Directory, Eric Clearinghouse, EBSCO, Economic Abstracts, Historical Abstracts, Index to Periodical Articles, Social Science Source, Social Science Index, Sociological Abstracts, and the University Reference System.

We wish to thank all authors for granting permission to publish their articles. We want to thank all those who reviewed these articles for publication.

National Social Science Association Journal
Volume 64 Issue #3
<u>Table of Contents</u>

Perceptions of Social and Emotional Learning Across K-12 Contexts

Janet Birkey
Eastern New Mexico University

Rebbecca Gossett
Eastern New Mexico University

Presley Shilling
Eastern New Mexico University

Tsitsi Nyabando
Eastern New Mexico University

Selenda Cumby
Eastern New Mexico University

Elizabeth Lester
Eastern New Mexico University

Abstract

Social-Emotional Learning (SEL) has been a part of the educational landscape for many years. However, the focus on SEL has grown since the COVID-19 Pandemic and global lockdown. While there is research on SEL, there is limited comparative research on the perceptions of principals, teachers, decision-makers, and implementers of SEL programs. This is a pilot study conducted in three elementary schools in rural southeastern New Mexico. The participants consisted of five principals and 44 teachers who completed a researcher-designed survey. The study found that both principals and teachers in the district's elementary schools viewed SEL as beneficial for students, particularly in self-management, social awareness, and relationship skills. In line with the literature review, both principals and teachers also had concerns about faculty training, funding, and scheduling. The principals in this survey perceived SEL to be moderately integrated into daily classroom instruction. The findings indicate that both principals and teachers value SEL practices and would like more instruction on delivery, funding to implement them with fidelity, and scheduling support to ensure consistent implementation.

Key Words: *social emotional learning (SEL), K-12 education*

Introduction

Social and emotional learning (SEL) is the process by which both adults and children acquire skills for healthy identities, emotion regulation, and goal achievement (New Mexico Public Education Department, 2021; Skoog-Hoffman et al., 2024). This definition supports a whole-child approach as it refers to the process through which individuals develop skills such as self-awareness, self-management, social awareness, relationship skills, and responsible decision-making (Collaborative for Academic, Social, and Emotional Learning [CASEL], 2020; De Luna, 2025; Kim et al., 2024). Life competencies are not taught in isolation, but rather as part of the overall school culture (Elias et al., 1997; Jones & Bouffard, 2012).

Although most educators are aware of Maslow's Hierarchy of Needs, which shows the need for a holistic approach, addressing those needs is often beyond the scope of the classroom instructor (Lawson et al., 2019). However, since the COVID-19 pandemic, the needs for safety, comfort, and security have become increasingly salient and inseparable from academics (Breaux, 2023; Skoog-Hoffman et al., 2024). Cognitive functioning is compromised when basic needs, such as students' breakfast, are not met (CASEL, n.d.; Dyson et al., 2023). It is difficult, at best, and often impossible, to teach a hungry or dysregulated student in the classroom (Dyson et al., 2023).

Mental health prevalence of anxiety and depression increased 25% during the first year of the COVID-19 pandemic (World Health Organization, 2022). In particular, young children and women were most affected (World Health Organization, 2022). Social isolation, loneliness, and family stress during the COVID-19 pandemic led to higher reports of increased clinginess, uncooperative behaviors, general misbehavior, increased inattention and impulsivity, anxiety, social isolation, and depression among children and adolescents (Meade, 2021). The Center on Reinventing Public Education (CRPE) found that approximately 30-40% of young people experienced increased impacts on their social-emotional and mental health (Hamilton & Gross, 2021). This large increase in concerns among this population raised alarms for countries to consider how to address these rising behavioral and mental health concerns within medical and K-12 school systems (Hamilton & Gross, 2021; Meade, 2021).

By February 2021, school system administrators and educators noted widespread declines in work, assignment completion, and social-emotional development (Hamilton & Gross, 2021). While data on mental health during this time was vast, data on social-emotional development were nearly nonexistent (Hamilton & Gross, 2021). Hamilton and Gross (2021) suggested that school districts move beyond individual SEL skills goals for students and instead focus on expanding SEL to support the environment rather than just individual goals. However, because data on student well-being and SEL outcomes are limited, it is important to expand measurement to better understand systemic changes and even barriers to social-emotional learning within the school district (Hamilton & Gross, 2021).

Social Emotional Learning Overview

SEL was first defined by Elias and colleagues (1997) as the process of acquiring competencies for managing emotions, setting goals, and maintaining relationships (Elias et al., 1997; Zins et al., 2007). In 2005, CASEL established a framework encompassing five core competencies: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making (CASEL, 2005; Durlak et al., 2011). SEL was a collaborative effort to understand the findings from various clinical, prevention, and youth development studies, which then moved into academic settings (Durlak et al., 2011; Greenberg et al., 2003).

SEL reflects varying perspectives across the academic landscape (Jones & Bouffard, 2012). SEL has experienced a decades-long shift of focus from primarily academic outcomes to 3 a broader emphasis on the "whole child," including social and emotional development (Aspen Institute, 2019; Hamilton et al., 2019). Becoming more comprehensive, faculty recognize that achievement is deeply intertwined with students' social and emotional development, a reality that has become increasingly apparent post-COVID-19 (Dyson et al., 2023; Skoog-Hoffman et al., 2024). Reportedly increasing rates of anxiety, depression, and behavioral challenges among students result in school personnel being more interested in SEL, seeing it as foundational to student success rather than a supplemental program (Breaux, 2023; Dyson et al., 2023).

The synthesized findings of a 2011 meta-analysis (Durlak et al., 2011) analyzed 213 universal, school-based SEL programs implemented from elementary through high school, totaling 270,034 students, with the goal of assessing the overall effectiveness of SEL programming on multiple student outcomes and to identify factors that moderate impact (Durlak et al., 2011). They found that SEL interventions demonstrated significant improvements across several domains: social-emotional skills; attitudes toward self and others; conduct and prosocial behavior; and academic performance, with an average gain equivalent to 11 percentile points in achievement (Durlak et al., 2011; Payton et al., 2008).

The CASEL Framework for SEL Implementation

With the establishment of SEL in the educational landscape, multiple frameworks have emerged to guide school districts in implementing it (Hamilton et al., 2019). One such framework, the CASEL framework, is widely used in schools as the basis for social-emotional learning programs in the United States (Bailey & Weiner, 2022; De Luna, 2025). The framework includes five components: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making (Bailey & Weiner, 2022; De Luna, 2025; Lawson et al., 2019). The originators of CASEL included experts in emotional intelligence and education who sought to advance the development of SEL for students in grades K-12 and to provide a basis for SEL programs to compare themselves against (De Luna, 2025; Jackson, 2012).

The CASEL framework includes a 45-item leader assessment that allows principals and teaching faculty to assess their personal SEL competencies and also provides tools to monitor the fidelity of their SEL programs (De Luna, 2025). Furthermore, the framework aligns with the concept of whole-child education, allowing leaders to ensure their instruction integrates academic and social-emotional development (De Luna, 2025; Hamilton et al., 2019). The CASEL framework also incorporates transformative SEL to support inclusiveness for diverse student populations (CASEL, 2023; Jagers et al., 2019).

Leadership Frameworks Supporting SEL Implementation: The Four Paths to Leadership

The CASEL framework is the starting point for SEL competencies in both K-12 and adult programs. However, the Four Paths of Leadership support school leaders in bridging the “knowing-doing gap” (Bailey & Weiner, 2022; Leithwood et al., 2017). CASEL defines the social and emotional learning competencies included in program implementation, but school leadership significantly shapes how those competencies are implemented (Bailey & Weiner, 2022; De Luna, 2025). One such leadership framework is the Four Paths to Leadership, which comprises the rational, organizational, family, and emotional paths (Leithwood et al., 2017). Taken together, these four components impact teaching faculty through the leadership in the school district and local schools (Bailey & Weiner, 2022). Focusing on more than purely academics, the Four Paths extend beyond instruction into the following areas: cognitive decision-making (rational), systems and structures (organizational), relational dynamics (family), and emotional intelligence and self-awareness (emotional) (Bailey & Weiner, 2022).

Principal Perceptions

Principal perceptions of SEL programs cannot be overlooked or undervalued as principals and school leaders play a critical role in shaping SEL implementation (Jones, 2016; Mahfouz et al., 2019; Skoog-Hoffman et al., 2024). Research shows that their beliefs and priorities directly influence school culture, teacher engagement, and resource allocation (Jones, 2016; Mahfouz et al., 2019; Oberle et al., 2016). Principals are the mandated gatekeepers of their schools and, as such, have significant decision-making authority, often resulting in varying levels of support for SEL programs (Jones, 2016; Hamilton et al., 2019). Van Velsor (2009) supported the belief that, while SEL is not as prominent for principals, school counselors can both advocate for and consult on it by working to ensure it is integrated with school needs and curriculum.

Overall, research shows that principals support the initiation and ongoing development of SEL programs in their schools (Hamilton et al., 2019; Mahfouz et al., 2019). This is generally considered a positive situation as SEL programs cannot be successful without principal endorsement and support (Jones, 2016; Oberle et al., 2016). While they may actively champion SEL in the early days of implementation, as the program continues, their view may shift toward viewing it as an added responsibility rather than a core component (Jones, 2016; Hamilton et al., 2019).

Principals often report fewer implementation challenges than teaching faculty, though they may actually face greater systemic hurdles (Hamilton et al., 2019). Key issues support the finding of a large “knowing-vs.-doing” gap among principals in the implementation of SEL (Bailey & Weiner, 2022; De Luna, 2025; Mahfouz et al., 2019). The challenges associated with these key issues include scheduling, community support for the program, and funding (Breaux, 2023; Hamilton et al., 2019; Lawson et al., 2019; Skoog-Hoffman et al., 2024). Despite challenges and concerns, principals recognize the pressing need for SEL programs in schools. Breaux found that principals reported that SEL was more needed post-pandemic than pre-pandemic (Breaux, 2023). Breaux’s findings showed that principals felt there was a greater deficit in skills of student self-management and relational skills due to factors like isolation, lack of structure, and trauma, which resulted in increased anxiety, aggression, and disengagement.

Research distinguishes between principals who are considered early adopters of SEL programs, stating that those who adopt early have more positive attitudes toward SEL, tend to be proactive planners, and allow greater teacher empowerment in SEL initiatives (Jones, 2016). The early adopters also extend their positivity to students by publicly modeling SEL concepts (e.g., greeting students in hallways and at buses) (Bailey & Weiner, 2022; Breaux, 2023). Principals who adopt later either have a neutral attitude (Jones, 2016) or consider SEL programs to be an unnecessary add-on to other responsibilities (Jones, 2016; Bailey & Weiner, 2022). Late adopters may also insist that teachers demonstrate that the SEL program is not working before considering any changes (Jones, 2016).

Teachers’ Perceptions

Research shows that teachers generally recognize the importance of SEL and value its role in fostering student development (Ee & Cheng, 2013; Hamilton et al., 2019; Karmina et al., 2024). However, their barriers differ from those of principals (Hamilton et al., 2019). The barriers often faced by teachers include implementation challenges, tension between SEL and the academic curriculum, and the perception that teaching academics and SEL concepts are competing priorities (Ee & Cheng, 2013; Lawson et al., 2019). Additionally, science, technology, engineering, and math (STEM) teachers found it more difficult to integrate SEL concepts into the curriculum compared to teachers of subjects such as language arts and history, who view the subject matter as a more "natural fit" (Ee & Cheng, 2013). Van Velsor (2009) further emphasized collaboration between school counselors and teachers through co-planning and co-delivering SEL lessons as a means of integrating SEL into the daily curriculum rather than treating it as a separate initiative .

While teachers tend to support SEL, they also report that time constraints, insufficient training, and limited resources impede their ability to fully implement SEL concepts (Breaux, 2023; Dyson et al., 2023; Hamilton et al., 2019). Instead, teachers may find themselves relying on their personal initiative to build these skills, often engaging in informal conversations or planning during their own time (Dyson et al., 2023; Karmina et al., 2024). Furthermore, teachers identify contextual factors, such as home-school disconnects, that further complicate implementation when school strategies conflict with or are not reinforced by students' home environments (Dyson et al., 2023).

Taken together, the literature demonstrates a strong consensus that SEL is essential for student success; however, implementation remains inconsistent (Cipriano et al., 2023; Hamilton et al., 2019; Jones & Bouffard, 2012). One of the most significant gaps lies in how SEL is enacted in practice, including not only the delivery of SEL lessons but also the availability of training, funding, and ongoing support necessary for effective implementation (Breaux, 2023; Cipriano et al., 2023; Ee & Cheng, 2013).

Institutional Barriers

The research literature on institutional barriers identifies lack of time, funding, and preparation as key hindrances in the development and implementation of SEL services (Hamilton et al., 2019; Skoog-Hoffman et al., 2024). Lack of time was identified as a barrier by 62% of respondents, who reported that SEL delivery competes with the time allotted for academic instruction (Skoog-Hoffman et al., 2024). Teachers facing a threshold for acceptable standardized test scores may be more inconsistent in SEL delivery, as academic subjects often take priority in the daily classroom schedule (Ee & Cheng, 2013; Hamilton et al., 2019). Furthermore, because SEL delivery is not always considered quantifiable, it can be sacrificed in favor of core content (Ee & Cheng, 2013; Jones, 2016). Van Velsor identified barriers such as academic pressures, time constraints, and the perception of SEL as optional, all of which contribute to inconsistent implementation within schools (Van Velsor, 2009).

Funding was identified as a barrier by 45% of respondents, leading to disparities in the quality of SEL delivery across school districts (Skoog-Hoffman et al., 2024). School counselors may be tasked with delivering SEL, but schools without counseling services or with high student-to-counselor ratios generally have teacher-led responsibility for it (Breaux, 2023; Hamilton et al., 2019). Finally, many teachers do not feel well-equipped to deliver SEL, which is often due to a lack of training (Dyson et al., 2023; Hamilton et al., 2019). In a national survey including 3,500 principals and 15,000 teachers, Hamilton et al. (2019) found that even though SEL is highly valued, teachers lack the time, training, and ancillary support to deliver SEL programs with fidelity.

Teachers often find they must teach themselves SEL concepts before they can teach them to students, and SEL is rarely addressed in preservice or in-service training (Dyson et al., 2023; Hamilton et al., 2019). This aspect of teacher-initiated SEL results in inconsistency of delivery, even within the same school building (Dyson et al., 2023; Skoog-Hoffman et al., 2024). Bowers et al.

(2017) also noted that school counselors are often underutilized in K–12 settings and frequently confined to administrative duties rather than leadership roles in SEL implementation.

Across the literature, research highlights that social-emotional learning (SEL) programs are an effective approach in both K–12 and adult-focused programs (Bailey & Weiner, 2022; Bowers et al., 2017; Cipriano et al., 2023; Durlak et al., 2011). However, findings indicate significant differences between principals' and teachers' perceptions regarding the importance and implementation of these programs (Hamilton et al., 2019; Skoog-Hoffman et al., 2024). Additionally, institutional barriers such as lack of time, insufficient funding, and inadequate professional preparation impact perceptions of effectiveness, hinder efforts to increase fidelity, and stall the comprehensive implementation of SEL services (Breaux, 2023; Dyson et al., 2023; Hamilton et al., 2019; Lawson et al., 2019; Skoog-Hoffman et al., 2024).

The research gap in the existing literature is that limited attention has been given to the perceptions of principals and teachers, those who make decisions, provide instruction, and incorporate SEL lessons, regarding how SEL is implemented, supported, and practiced in schools (Breaux, 2023; De Luna, 2025; Jones & Cater, 2020). While multiple studies have examined various aspects of SEL, few have compared principals and teachers within a single research design (Hamilton et al., 2019). This underscores the need for comparative studies that examine principal and teacher perspectives simultaneously to ensure schoolwide alignment and communication (Hamilton et al., 2019; Skoog-Hoffman et al., 2024).

The research questions that frame this pilot study were:

1. How do New Mexico schools implement the state's SEL Framework components in daily instruction and school culture?
2. How do the principal perceptions regarding SEL implementation on their campus compare to the teacher perceptions?

Methodology

Research Design

This quantitative study serves as a preliminary examination of how principals and teachers perceive SEL, how it is implemented in daily instruction and school culture in New Mexico's schools, and its perceived effects on students. The researchers developed two surveys (one for principals and one for teachers) for a larger, future mixed-methods statewide study. The results reported in this manuscript are from surveys that were piloted for validity and reliability across three elementary campuses in a single rural school district. The piloting was also meant to help identify any problems with the wording of the questions or other aspects, such as sentence length or the order of some items. A survey was appropriate for understanding principals' and teachers' perceptions of SEL implementation and how schools implemented the state's SEL Framework.

Participants

The participants were five principals and 44 elementary school teachers from a small rural school district. Participants for this pilot study were selected using a representative sampling approach for the larger study. Statistics indicate that New Mexico districts have more female (64%) than male (36%) principals; 58% of districts are considered rural; 61% have >70% of students economically disadvantaged; and 18.8% have English Learners. The district selected for the elementary school pilot included 100% female principals, 100% rural, 76% of campuses with>70% economically disadvantaged students, and 9.8% English Learners. This demonstrates how the sample utilized for the pilot was similar to the population of New Mexico. Otherwise, the demographics will vary by age and educational experience.

The pilot study used a single rural district with three elementary school campuses. The participants included principals and assistant principals at the three elementary schools, as well as all teachers of record at those schools.

Survey Instrument

The instrument used in this study comprised two surveys, one for principals and the other for teachers. The survey items were based on available research on the topic. The face and content validity were reviewed by colleagues, who provided feedback on the survey's wording, emphasizing precision and clarity. Their suggested changes were made before administering the survey. The surveys utilized a 5-point Likert rating scale that contained demographic questions along with questions that corresponded to five objectives: 1) How schools implement New Mexico's SEL Framework components in daily instruction and school culture; 2) Perceived effects of SEL programs on students' social-emotional competencies; 3) How trauma-informed care and restorative justice complement SEL; 4) How districts select and adapt SEL programs to meet community and student needs; and 5) Challenges and successes in integrating SEL with academics and broader school supports. These two surveys were developed and submitted to the university IRB with the rest of the application. Both surveys were measured for reliability and validity. Validity was established through expert review by five faculty members, who examined the items' alignment with the constructs under investigation.

The pilot study aimed to evaluate the instrument's validity, reliability, and clarity for participants. This was a preliminary step for a larger study that the researchers planned to conduct with middle school social studies teachers and principals throughout New Mexico. The researchers also aimed to use the research findings to develop interview questions for the larger mixed-methods study.

Data Collection Procedures

After IRB approval for a larger-scale mixed-methods study, the newly developed surveys were piloted to assess reliability and validity. It was decided that this pilot study would complete this task through quantitative measurement utilizing the two surveys. As previously discussed, a district was selected to closely mirror New Mexico's population. Instead of surveying middle school principals and teachers in the large-scale study, the researchers used the pilot survey of elementary principals and teachers. The superintendent was contacted for approval to conduct the surveys at the district’s three elementary campuses. Further consent of the district was not required. The superintendent recommended contacting all potential participants during the district’s staff development day, as time to participate was built into the agenda.

The surveys were accessed by all principals, assistant principals, and teachers of record through a Qualtrics link on either a computer or phone. The opportunity to complete the surveys was presented to principals and teachers, and they were given time to volunteer or not. There were 44 teachers and five principals who completed the full survey.

Ethical Considerations

Approval for the study was gained from the researcher’s university IRB in 2025. The surveys were set up in Qualtrics to allow all participants to choose whether to participate through a consent form approved by the university’s IRB. The Qualtrics surveys included a consent question asking whether participants would participate. If the participant selected to accept the consent, the survey commenced. If the person did not wish to participate, they selected "do not accept" and were exited from the survey. All data collected were confidential, with no identifiable descriptors, and stored in a password-protected project. The participants completed a survey that did not identify their specific campus and will not include any identifiable descriptors.

Data Analysis

After completing the surveys, the researchers manually standardized the collected data to ensure consistency. Survey responses were reviewed for completeness and accuracy. Researchers then identified comparable questions between the two data sets (principals and teachers). All Likert-type questions were recorded into a unified scale of 1 to 5, and a clean data dictionary was created. The 1 represented Very Low / Strongly Disagree, while the 5 represented Very High / Strongly Agree. The responses were then imported into Excel, and the questions were aligned appropriately across the two survey datasets. Blank answers were not recorded and remained blank.

It is also essential to examine normality, linearity, and homoscedasticity to assess the adequacy of the data (Mertler et al., 2022). The researchers then used the cleaned data to compute descriptive statistics, including frequencies, percentages, means, and standard deviations, to summarize the survey responses. These statistics were used to provide an overall description of the responses and response patterns (Mphaphuli et al., 2024). Measures of central tendency, variability, relative position, and relationship were evaluated (Mertler et al., 2022).

The statistical analyses selected to address the research questions included the Mann–Whitney U test, the Chi-Square Test of Independence, and Cramér's V. Mann–Whitney U Test was run across all matched columns. A Chi-square test of independence was also conducted to examine differences in response distributions between the two groups. To add a much deeper layer of interpretation to the findings, a Cramér's V test was conducted. Using Cramér's V would strengthen interpretation by allowing the analysis to move beyond whether there is a difference to how meaningful that difference may be. The Mann–Whitney U test was used to determine whether there were statistically significant differences between the two independent groups, principals and teachers. Because the survey items were measured on Likert-type scales, the data were treated as ordinal rather than continuous. As a result, a nonparametric test was more appropriate than a parametric alternative. Additionally, the Mann–Whitney U test was selected because it is robust for comparing groups with unequal sample sizes, as in the principal and teacher participant groups in this study. This test evaluated differences in the distribution of ranked responses between the two groups.

The Chi-Square Test of Independence was also employed as a nonparametric statistical procedure to examine whether an association existed between participant group membership and survey response categories. Similar to the rationale for the Mann–Whitney U test, the Likert-scale survey items provided categorical responses rather than continuous numerical measurements, and the study involved comparisons between two independent groups. The Chi-Square Test of Independence complemented the Mann–Whitney U analysis by evaluating patterns in response frequencies across categories. For both statistical tests, the alpha level for determining statistical significance was set at $p < .05$.

Cramér's V was selected as a follow-up measure to the Chi-Square Test of Independence to assess the strength of association between participant group (principals and teachers) and the distributions of survey responses. Although the Chi-Square Test identifies whether a statistically significant relationship exists, it does not provide information regarding the magnitude or practical significance of that relationship. Because the variables analyzed were categorical, Cramér's V was an appropriate measure of effect size. The resulting effect size values enabled interpretation of the strength and practical significance of the observed group differences.

Results

Principal Perceptions

The research question examined how New Mexico schools implement the state's SEL Framework components in daily instruction and school culture, as well as principals' perceptions of SEL implementation on their campuses. Descriptive statistics, including means and standard deviations for the principals, are shown in Table 1.

Table 1
Principal Survey Descriptive Statistics

No	Item	Mean	SD
Q1	The state's SEL Framework components are clearly integrated into daily instruction across my school.	3.20	0.84
Q2	SEL practices guided by the state framework are evident in the overall school culture.	3.20	0.84
Q3	Staff across all roles (teachers, education assistants, support staff, administrators) consistently model SEL competencies.	3.40	0.89
Q4.1	Teachers receive adequate training to implement the state's SEL Framework effectively.	1.80	1.30
Q5.1	Staff across all roles (education assistants, support staff, administrators) receive sufficient training to effectively implement the state's SEL Framework.	1.80	0.84
Q6.1	School leadership consistently supports the use of the SEL Framework across classrooms.	2.80	1.30
Q7	SEL activities aligned with the state framework are regularly promoted during non-instructional times (e.g., recess, transitions).	3.00	0.71
Q1.1	SEL programs in my school have improved students' self-awareness.	3.80	0.45
Q2.1	SEL programs have helped students better manage their emotions and behaviors (self-management).	4.00	0.71
Q3.1	SEL initiatives have increased students' social awareness and empathy towards others.	4.00	0.71
Q4.2	SEL has strengthened students' relationship skills, including cooperation and communication.	4.00	0.71
Q5.2	Students demonstrate improved responsible decision-making as a result of SEL programs.	4.00	0.71
Q1.2	I am aware of the trauma-informed care practices being integrated into the SEL framework used at my school or by the state.	2.60	1.34

Q2.2	Trauma-informed care practices effectively support students' social and emotional needs alongside SEL.	4.00	1.00
Q3.2	Restorative justice practices are integrated with SEL to address student behavior and relationships	3.60	0.55
Q4.3	Staff feel confident applying trauma-informed strategies together with SEL instruction.	2.80	1.30
Q5.3	Restorative justice approaches reinforce SEL skills in conflict resolution.	3.40	1.52
Q6.2	The combination of trauma-informed care, restorative justice, and SEL creates a supportive learning environment.	4.20	0.84
Q1.3	SEL programs are selected based on the specific needs of our school community and students.	2.40	0.89
Q2.3	The SEL program used by my school district effectively meets the needs of both the community and its students.	3.20	1.48
Q3.3	SEL programs are adapted to fit the cultural and demographic characteristics of our students.	3.60	0.55
Q4.4	Teachers and staff are involved in decisions about SEL program selection and implementation.	2.60	1.14
Q5.4	Our school provides sufficient resources to support SEL program adaptation and implementation.	2.20	0.84
Q6.3	SEL programs are regularly reviewed and updated to improve their effectiveness.	2.20	0.84
Q1.4	Integrating SEL with academic instruction presents significant challenges in my school	3.00	0.71
Q3.4	Collaboration between SEL staff and academic teachers is effective in my school.	3.00	0.71
Q4.5	Time constraints limit the successful integration of SEL with other school supports.	4.00	0.71
Q5.5	Our school has experienced clear successes in embedding SEL across academic and support services.	3.00	1.22
Q16	How would you rate the overall effectiveness of these SEL programs in improving student social-emotional skills?	3.40	0.89
Q17	How well do these programs support positive school climate, classroom management, and student-teacher relationships?	3.40	0.55
Q18	What challenges have been encountered in implementing these SEL programs at your school?	3.40	0.55
Q19	What challenges have been encountered in implementing these SEL programs at your school?	3.20	0.84

Note. Mean and standard deviations on a 5-point scale. n=5

The principals reported varying levels of familiarity with the New Mexico SEL Framework. Forty percent reported being "slightly familiar," another forty percent indicated they were "moderately familiar," and twenty percent stated they were "not familiar at all." Only one principal mentioned that their school has a designated SEL coordinator or team. While the school leaders felt that SEL practices are moderately evident in the school culture, they identified a significant training gap. They noted that components of SEL are somewhat integrated into daily instruction, but both teachers and support staff receive relatively limited training to implement the framework effectively. They also reported a resource challenge and a review of the effectiveness or appropriateness of SEL for the school.

Despite this gap, the participants reported strong positive impacts on student outcomes, with ratings for various skills as follows: Self-Awareness: 4.00, Self-Management: 4.40, Social Awareness/Empathy: 4.40, and Relationship Skills: 4.40. The principals indicated that the most frequently implemented program is "Capturing Kids' Hearts" (CKH), followed by "Second Step" and "Trauma-Informed Training." They rated these programs as very effective in improving students' social-emotional skills.

Teacher Perceptions

The research question examined how New Mexico schools implement the state's SEL Framework components in daily instruction and school culture, and how teachers perceive the SEL implementation on their campuses. Descriptive statistics for the teachers, including means and standard deviations, are shown in Table 2.

Teachers reported diverse levels of familiarity. Approximately 22.7% identified as highly familiar (level 5), while 34% reported moderate familiarity (level 3). Notably, 18% reported having very little familiarity (level 1). Most respondents (55.8%) have received formal SEL training, while 44.2% have not. Teachers generally perceive a positive alignment with the SEL framework in their daily work. Teachers agree that the framework components are clearly integrated into daily classroom instruction and that SEL practices are reflected in the overall school culture. They also reported that staff across all roles are seen as consistently modeling SEL competencies. Teachers feel they receive adequate training to implement the state framework effectively, though this score is slightly lower than the integration scores, suggesting a need for more support.

Table 2

Teacher Survey Descriptive Statistics (5-Point Likert Scale)

No	Item	N	Mean	SD
Q4	How familiar are you with the New Mexico Social Emotional Learning (SEL) Framework?	44	3.34	1.35
Q1	The state's SEL Framework components are clearly integrated into daily classroom instruction.	44	3.70	1.07
Q2	SEL practices guided by the state framework are reflected in the overall school culture.	44	3.70	1.02
Q3	Staff across all roles (teachers, education assistants, support staff, administrators) consistently model SEL competencies.	44	3.61	1.02
Q4.1	I receive sufficient training to effectively implement the state's SEL Framework in my teaching.	28	3.43	0.84
Q5.1	Staff across all roles (teachers, support staff, administrators) receive sufficient training to effectively implement the state's SEL Framework.	33	3.39	1.00
Q6.1	The school leadership supports consistent use of the SEL Framework across all classrooms.	44	3.57	1.19
Q7	SEL activities aligned with the state framework are regularly incorporated during non-instructional times (e.g., recess, transitions).	44	3.50	1.05
Q1.1	SEL programs in my school have improved students' self-awareness	41	3.59	1.07
Q2.1	SEL programs have helped students better manage their emotions and behaviors (self-management).	41	3.59	0.95
Q3.1	SEL initiatives have increased students' social awareness and empathy towards others.	38	3.59	0.94
Q4.2	SEL has strengthened students' relationship skills, including cooperation and communication.	40	3.58	0.90
Q5.2	Students demonstrate improved responsible decision-making as a result of SEL programs.	41	3.44	0.92
Q5.3	Restorative justice approaches help reinforce SEL skills in conflict resolution.	39	3.38	0.99
Q1.2	I am aware of the trauma-informed care practices being integrated into the SEL framework used at my school or by the state.	99	3.41	1.12
Q2.2	Trauma-informed care practices in my school effectively support students' social and emotional needs alongside SEL.	39	3.44	0.97
Q3.2	Restorative justice practices are integrated with SEL to address student behavior and relationships.	38	3.32	1.07
Q4.3	I feel confident applying trauma-informed strategies together with SEL instruction	38	3.32	1.14
Q6.2	The combination of trauma-informed care, restorative justice, and SEL creates a more supportive learning environment.	39	3.59	1.04
Q1.3	My district selects SEL programs based on the specific needs of our school community.	36	3.56	0.91
Q2.3	The SEL program used by my school district effectively meets the needs of both the community and its students.	36	3.39	0.96
Q3.3	SEL programs are adapted as needed to fit the cultural and demographic characteristics of our students.	36	3.65	1.07

Q4.4	Teachers are involved in decisions about which SEL programs are implemented.	37	3.05	1.22
Q5.4	The district provides sufficient resources to support SEL program adaptation and implementation	37	3.27	1.12
Q6.3	Our district regularly reviews and updates SEL programs to improve their effectiveness.	36	3.25	1.05
Q1.4	Integrating SEL with academic instruction presents significant challenges in my school.	37	3.19	0.88
Q2.4	SEL integration has positively impacted students' academic engagement and performance.	36	3.58	0.87
Q3.4	Collaboration between SEL staff and academic teachers is effective in my school.	36	3.58	0.87
Q4.5	Time constraints limit the successful integration of SEL with other school supports.	37	3.54	0.90
Q5.5	Our school has experienced clear successes in embedding SEL across academic and support services.	37	3.51	0.90
Q16	How would you rate the overall effectiveness of these SEL programs in improving student social-emotional skills?	36	3.17	1.03
Q17	How well do these programs support positive school climate, classroom management, and student-teacher relationships?	36	3.42	0.87
Q18	What challenges have been encountered in implementing these SEL programs at your school?	34	3.47	0.61
Q19	How well do these SEL programs meet the diverse needs of students at your school?	36	3.25	0.94

Perceived Impact on Students and Program Effectiveness

Teachers observed consistent benefits for students across the following social-emotional domains: Self-Awareness (3.59), Self-Management (3.70), Social Awareness & Empathy (3.61), Relationship Skills (3.58), and Responsible Decision-Making (3.44). Teachers reported that Capturing Kids' Hearts (CKH) is the dominant SEL program implemented in these schools, mentioned by nearly every respondent who specified a program (n=36). On average, teachers rated these programs as "Moderately effective" in improving students' social-emotional skills.

Principal Perceptions Regarding SEL Implementation on their Campus Compared to the Teacher Perceptions

The second research question examined how principals' perceptions of SEL implementation on their campuses compare with those of teachers. Tables 3, 4, 5, and Figures 1 and 2 show statistics comparing the two. While both groups rated the impact of SEL programs positively, teachers reported slightly lower mean scores for "Sufficient Training" than for "Framework Integration." This aligns with the principal data, which also highlighted training as a lower-scoring area, indicating a consistent statewide or district-wide opportunity for professional development. The most significant finding is that teachers generally rated certain aspects of SEL implementation higher than principals did, particularly in program selection and needs alignment. There is a significant difference (-1.16).

Teachers felt more strongly than principals that SEL programs are selected based on the specific needs of the school community and students. This suggests that principals may be more critical of the selection process or more aware of misalignments than their teaching staff. Teachers reported higher awareness/confidence in trauma-informed care practices compared to principals, with a difference of -0.81. Teachers rated the integration of SEL into daily instruction and school culture slightly higher than principals. Conversely, principals rated overall effectiveness higher than teachers. While principals see the "big picture" impact as more successful, teachers (who implement the programs daily) are slightly more conservative in their rating.

Table 3

Summary Comparison Between Principal and Teacher Perceptions

Item	Topic	Principal Mean	Teacher Mean	Difference (P-T)
Q1.3	Program Selection needs	2.40	3.56	-1.16*
Q1.2	Trauma-Inforced Awareness	2.60	3.41	-0.81
Q1	Framework Integration	3.20	3.70	-0.50
Q16	Overall Effectiveness	3.40	3.17	+0.23
Q1.1	Self-Awareness Impact	3.80	3.59	+0.21

**Significant difference ($p > 0.05$).*

Figure 1

Boxplot Comparison (All Questions Combined)

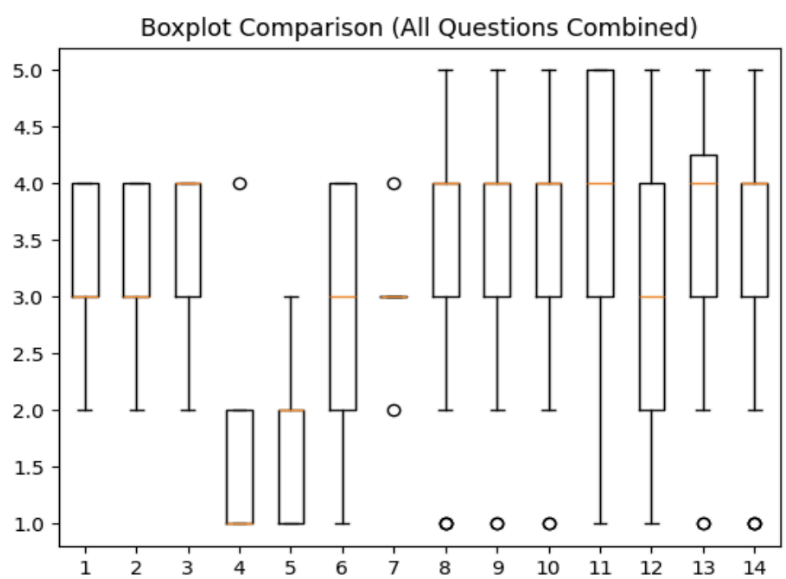


Table 4

Summary Comparison Between Principal and Teacher Perceptions

Question	χ^2 (Chi2)	p-value	df
Q1	3.94	0.415	4
Q2	2.78	0.595	4
Q3	1.70	0.791	4
Q4	9.73	*0.045	4
Q5	4.86	0.302	4
Q6	3.06	0.547	4
Q7	3.84	0.427	4

**The distribution of responses differs between groups*

As shown in Table 5 below, there were no significant differences in responses for the questions that covered the objective. These were questions about the perceived effects of SEL programs on students' self-awareness, self-management, social awareness, relationship skills, and responsible decision-making.

Table 5
Perceived Effects of SEL Programs on Students' Self-Awareness, Self-Management, Social Awareness, Relationship Skills, And Responsible Decision-Making

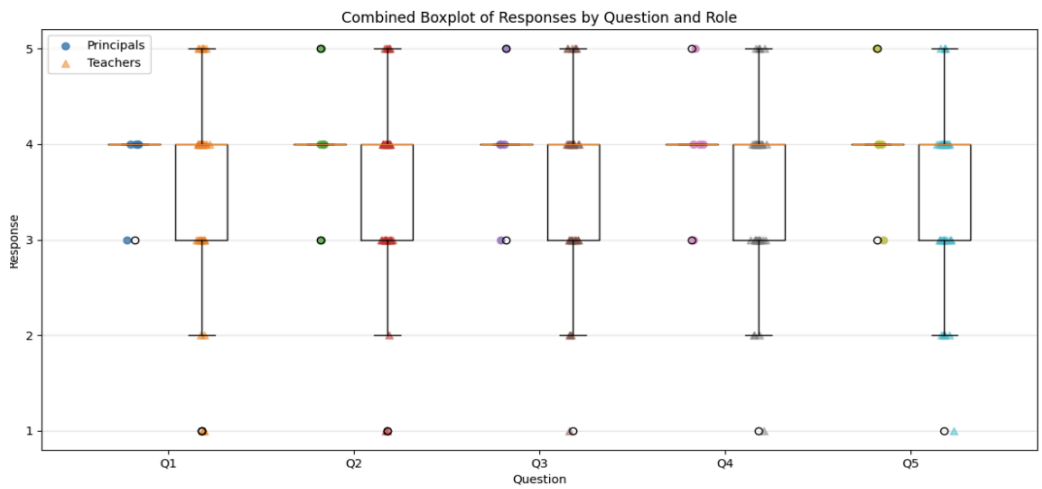
Question	Principals	Teachers	U	p-value
Q1	4.0	4.0	110.5	0.777
Q2	4.0	4.0	128.0	0.345
Q3	4.0	4.0	121.5	0.355
Q4	4.0	4.0	126.0	0.322
Q5	4.0	4.0	137.5	0.195

Table 6
Results for Objective 3

Question	Chi-square	df	p-value	Cramer's V
Q1	2.757	4	0.599	0.245
Q2	1.155	4	0.885	0.158
Q3	1.189	4	0.880	0.164

Q4	1.125	4	0.890	0.158
Q5	1.788	4	0.775	0.197

Figure 2
Combined Boxplot of Responses by Question and Role



Discussion

While teachers believe the programs are better "fitted" to their students' needs than principals do, they also report slightly lower overall effectiveness scores. This could indicate that although staff perceive the program choice positively, the execution or results still have room for improvement. Several numerical differences exist between principal and teacher perceptions. Although most are not statistically significant, likely due to the small principal sample size (n=5), they remain relevant because they highlight how perceptions can differ between school administrators and teachers. For SEL to be truly effective and for students to gain the maximum benefit, it's crucial that teachers, administrators, and staff are aligned.

Whatever the setting, there remains a consistent concern about the lack of formal training and professional development to adequately prepare educators, which is consistent with previous research (Dyson et al., 2023; Hamilton et al., 2019; Mahfouz et al., 2019). Instructor capacity and SEL buy-in represent additional implementation-level factors that further influence program success (Lawson et al., 2019; Skoog-Hoffman et al., 2024). Collectively, these barriers contribute to the uneven landscape of SEL across school districts and may help explain the differences in principal and teacher perspectives (Hamilton et al., 2019). In this study, however, the results showed teacher SEL buy-in. While evidence-based SEL programs can yield positive outcomes, their effectiveness is often limited by inconsistent implementation (Cipriano et al., 2023). This calls for further exploration of the effectiveness with a larger sample. Teacher and principal perceptions and buy-in are important, as a more sustainable model that integrates SEL into everyday interactions, routines, and instructional practices is more scalable, practical, and aligned with academics (Jones & Bouffard, 2012). Further research is needed on funding considerations, implementation processes, faculty training, and institutional (including community) support in educational settings (Breaux, 2023; De Luna, 2025; Hamilton et al., 2019).

Limitations and Conclusions

The main limitation of this study, which comprised three elementary schools with teaching faculty and principals, is its small sample size, which may limit its generalizability to other contexts. As mentioned earlier, these results represent a pilot assessment; hence the small sample size (n=5). The researchers feel that the data gathered cannot be generalized to other school districts because the study included only faculty from a Title I rural middle school.

This pilot study examined the comparative perceptions of principals and teachers regarding the implementation of Social-Emotional Learning in a single-district, rural New Mexico school and classrooms. Most principals' and teachers' perceptions in this study align with prior research that SEL is valuable, but also agree that it takes intentional effort to deliver an SEL program with fidelity and consistency. This also confirms the "knowing v. doing gap". This study reinforces current data on SEL programs and contributes to the research literature on rural schools and the challenges they face. The classroom plays an integral role in students' lives, and instructors are now tasked with delivering services that were previously considered outside their scope of education and classroom instruction. However, as the research literature confirms, the need for SEL programs has never been greater. While both principals and teachers have concerns about SEL, these two groups must find common ground if SEL is to become a part of their school culture and subsequently extend into the community. To successfully implement SEL programs, school and community support must be in place, professional training for both principals and teachers must be provided, and funding must be secured.

Future research could focus on larger sample sizes and extending grade levels beyond elementary grades. Additionally, a validated survey instrument could be implemented.

References

Aspen Institute, National Commission on Social, Emotional, & Academic Development (NCSEAD). (2019). From a nation at risk to a nation at hope: Recommendations from the National Commission on Social, Emotional, and Academic Development. *ERIC.*

https://eric.ed.gov/?q=source%3A%22Aspen+Institute%22&ff1=locDistrict+of+Columbia&id=ED606337

Bailey, J., & Weiner, R. (2022). Interpreting social-emotional learning: How school leaders make sense of SEL skills for themselves and others. *School Leadership Review, 16*(2), 1–34.

Bowers, H., & Lemberger, M. E. (2017). A person-centered humanistic approach to performing evidence-based school counseling research. *Person-Centered & Experiential Psychotherapies, 15*(1), 55–66. https://doi.org/10.1080/14779757.2016.1139502

Breaux, T. C. (2023). Perceptions of elementary school principals on the social emotional learning needs of students, aligned to the CASEL framework, during the COVID-19 pandemic (Doctoral dissertation, Virginia Tech). Virginia Tech Digital Repository. https://vtechworks.lib.vt.edu/items/6e7a75fa-cb98-486b-9777-5e2cd89e9aa7

Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development, 94*(5), 1181–1204. https://doi.org/10.1111/cdev.13968

Collaborative for Academic, Social, and Emotional Learning. (2005). Safe and sound: An educational leader's guide to evidence-based social and emotional learning (SEL) programs (Illinois ed.). https://casel.org/safe-and-sound-guide-to-sel-programs/

Collaborative for Academic, Social, and Emotional Learning. (2020). CASEL's SEL framework: What are the core competence areas and where are they promoted? CASEL. https://casel.org/casel-sel-framework-11-2020/

Collaborative for Academic, Social, and Emotional Learning. (2023). What is the CASEL framework? CASEL. https://casel.org/fundamentals-of-sel/what-is-the-casel-framework/

Collaborative for Academic, Social, and Emotional Learning. (n.d.). Fundamentals of SEL. CASEL. https://casel.org/fundamentals-of-sel/

De Luna, K. (2025). K–12 leaders and their self-perceptions of social emotional learning competence [Doctoral dissertation, Abilene Christian University]. Digital Commons @ ACU. https://digitalcommons.acu.edu/etd/856/

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Dyson, B., Shen, Y., Howley, D., & Baek, S. (2023). Social emotional learning matters: Interpreting educators' perspectives at a high-needs rural elementary school. *Frontiers in Education, 8*, 1–10. https://doi.org/10.3389/feduc.2023.1100667

Ee, J., & Cheng, Q. L. (2013). Teachers' perceptions of students' social-emotional learning and their infusion of SEL. *Journal of Teaching and Teacher Education, 1*(2), 59–72. https://doi.org/10.12785/jtte/010201

Elias, M. J., Zins, J. E., Weissberg, R. P., Frey, K. S., Greenberg, M. T., Haynes, N. M., Kessler, R., Schwab-Stone, M. E., & Shriver, T. P. (1997). Promoting social and emotional learning: Guidelines for educators. ASCD.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*(6-7), 466–474. https://doi.org/10.1037/0003-066X.58.6-7.466

Hamilton, L. S., Doss, C. J., & Steiner, E. D. (2019). Teacher and principal perspectives on social and emotional learning in America's schools: Findings from the American Educator Panels (RR-2991-BMGF). RAND Corporation. https://doi.org/10.7249/RR2991

Hamilton, L., & Gross, B. (2021). How has the pandemic affected students' social-emotional well-being? A review of the evidence to date. Center on Reinventing Public Education. https://www.crpe.org/publications/how-has-the-pandemic-affected-students-socialemotional-well-being-a-review-of-the-evidence-to-date/

Jackson, J. (2012). Partnering to bring social and emotional learning to school [Commentary]. *Social Policy Report, 26*(4), 26–27.

Jagers, R. J., Rivas-Drake, D., & Williams, B. (2019). Transformative social and emotional learning (SEL): Toward SEL in service of educational equity and excellence. *Educational Psychologist, 54*(3), 162–184. https://doi.org/10.1080/00461520.2019.1623032

Jones, S. M., & Bouffard, S. M. (2012). Social and emotional learning in schools: From programs to strategies and commentaries. *Social Policy Report, 26*(4), 1–33. https://doi.org/10.1002/j.2379-3988.2012.tb00073.x

Karmina, S., Dyson, B., & Setyowati, L. (2024). Teachers' perspectives on implementing cooperative learning to promote social and emotional learning. *Jurnal Cakrawala Pendidikan, 43*(2), 470–479. https://doi.org/10.21831/cp.v43i2.68447

Kim, E. K., Allen, J. P., & Jimerson, S. R. (2024). Supporting student Social Emotional Learning and development. *School Psychology Review, 53*(3), 201–207. https://doi.org/10.1080/2372966X.2024.2346443

Lawson, G. M., McKenzie, M. E., Becker, K. D., Selby, L., & Hoover, S. A. (2019). The core components of evidence-based social emotional learning programs. *Prevention Science, 20*(4), 457–467.

Leithwood, K., Sun, J., & Schumacker, R, (2017). How school leadership influences student learning: A test of "the four paths model". *Educational Administration Quarterly, 56*(4), 570–599. https://doi.org/10.1177/0013161X19878772

Mahfouz, J., Greenberg, M. T., & Rodriguez, A. (2019). *Principals' social and emotional competence: A key factor for creating caring schools* [Issue Brief]. Pennsylvania State University.

Meade J. (2021). Mental health effects of the COVID-19 pandemic on children and adolescents: A review of the current research. *Pediatric Clinics of North America, 68*(5), 945–959. https://doi.org/10.1016/j.pcl.2021.05.003

Mphaphuli, L. M. E., Coetzee, S. K., Tau, B., & Ellis, S. M. (2024). Nursing categories' perceptions of the practice environment and quality of care in North West Province: a cross-sectional survey design. *BMC Nursing, 23*(1), 1–12. https://doiorg.adams.idm.oclc.org/10.1186/s12912-024-01998-7

New Mexico Public Education Department. (2021). New Mexico Public Education Department's Social Emotional Learning Framework. https://web.ped.nm.gov/wpcontent/uploads/2025/01/SHSB_NM.SEL_.Framework.document.1.7.21-002.pdf

Oberle, E., Domitrovich, C. E., Meyers, D. C., & Weissberg, R. P. (2016). Establishing systemic social and emotional learning approaches in schools: A framework for school-wide implementation. *Cambridge Journal of Education, 46*(3), 277–297. https://doi.org/10.1080/0305764x.2015.1125450

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews* [Technical Report]. Collaborative for Academic, Social, and Emotional Learning. https://files.eric.ed.gov/fulltext/ED505370.pdf

Skoog-Hoffman, A., Miller, A. A., Plate, R. C., Meyers, D. C., Tucker, A. S., Meyers, G., Diliberti, M. K., Schwartz, H. L., Kuhfeld, M., Jagers, R. J., Steele, L., & Schlund, J. (2024). Social and Emotional Learning in U.S. schools: Findings from CASEL's nationwide policy scan and the American Teacher Panel and American School Leader Panel Survey. *Rand*. https://www.rand.org/pubs/research_reports/RRA1822-2.html

Van Velsor, P. (2009). School counselors as social-emotional learning consultants: Where do we begin? *Professional School Counseling, 13*(1), 50–58. https://doi.org/10.1177/2156759X0901300106

World Health Organization. (2022, March 2). COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2007). The scientific base linking social and emotional learning to school success. *Journal of Educational and Psychological Consultation*, *17*(2-3), 191–210.

Six Minute Solution on Student Reading Fluency in the Third Grade

Megan Schneck
Graduate Student Austin Peay State University

Abstract

This study investigates how the implementation of the Six Minute Solution affects the reading fluency, as measured by the word count per minute (WCPM), of third grade students over the course of the 2025-2026 school year. This program uses short passages which the students read with a leveled partner each day for a week and together track their reading fluency and growth as related to individualized reading goals. The purpose of this study is to examine the effectiveness of the Six Minute Solution (Adams & Brown, 2007) as a best instructional practice for improving the reading fluency and rate of all students by investigating the implementation of this program in a self-contained third grade classroom in Middle Tennessee. Through the course of this study, 14 of the 17 students experienced growth in their reading fluency while the remaining three students had a decrease in their WCPM from fall to spring screening. While not all students exhibited the same growth in their reading rate, there was an increased enjoyment found in reading for all students who took part in this study. The Six Minute Solution is now being implemented in additional classrooms throughout the Middle Tennessee area to help foster enjoyment for reading and help students improve their reading fluency and rate through daily repeated readings.

Key Words: Fluency, Repeated Reading, Six Minute Solution, WCPM, Third Grade

Introduction

By the 3rd grade, students experience a significant instructional shift—from the "learning to read" focus of kindergarten through 2nd grade to the "reading to learn" expectations of the upper grade levels. This shift in expectations can cause even the strongest readers to stumble over their new responsibilities and rigor related to their reading abilities. While whole group instruction is valuable to teach students the skills and strategies which they will need to be successful readers, there is not often time to simply practice reading to build student reading fluency. All parts of reading instruction are vitally important, especially in the third grade as this is where the TN State Law on retention takes hold. According to Brown and Mumpower (2023), “T.C.A. 49-6-3115 requires any 3rd grade student who does not achieve grade-level proficiency on the ELA portion of their [Tennessee Comprehensive Assessment Program] TCAP exam to be retained, unless the student can meet specific criteria.” The TCAP assessment has always been rigorous and now, with this added pressure of withholding students based on their reading proficiency skills, there is an even greater weight put on the scores which are received for both teachers and students.

This test is highly involved and placed solely on the responsibility of the student who can receive no help from the teachers when it comes to reading questions, answer choices, or passages. The academic level of the texts is set to grade level appropriateness but there are many students who come not yet reading at grade-level standard (The Nation's Report Card, 2022). This shows the vital importance of building the reading fluency and stamina of third grade students long before they encounter this end of year testing which holds such heavy weight. Before the state retention law was put in place, 64 percent of third grade students would have been considered for retention during the 2021-2022 school year based on the Tennessee Learning Loss Remediation and Student Acceleration Act (T.C.A. 49-6-3115) which went into effect the following school year (Brown & Mumpower, 2023). This number does not reflect the number of students who would have actually been retained as there are interventions put in place to address this concern – such as attending summer school, tutoring, or an exemption based on documented disability, lack of English language knowledge, or having previously been retained. Even though there are interventions in place to help push students forward, a lack of reading fluency is something that keeps students from being able to achieve their best in any grade level.

The National Center on Improving Literacy (2020, p. 1) states, “fluency is not an end in itself but a critical gateway to comprehension. Fluent reading frees cognitive resources to process meaning.” If students are still struggling to fluently read the words that they are seeing, they do not have the mental space to focus on making meaning of what they have read. This is vitally important as they are being tested on what they read, not *if* they can read it. To help build stamina for comprehension, we must first build more fluent readers as demonstrated with Scarborough’s Rope which shows how language comprehension, or the ability to understand what is read, and word recognition, the act of reading to include both fluency and rate, are intertwined for skilled reading and meaning making (Scarborough, et al., 2009). When a student struggles in one area of the reading rope, it affects the strength of the rest of the intertwined pieces.

One way to build reading fluency that is supported by researchers such as Rasinski, Samuels, and Shanahan, is using repeated readings (Rasinski, 2024; Samuels, 1997; Shanahan, 2017). Repeated reading is a literacy strategy where students read the same short text passage multiple times to increase their fluency and comprehension by becoming more familiar with the text and increasing their ability to read naturally with fluency and expression. As the students grow more familiar with what words they are reading, they are able to place more focus and cognitive energy on comprehending the meaning and information in the text. Rasinski (2024) states that he saw a great shift in the reading fluency of his own students after beginning to implement repeated readings as the struggling readers saw that they could be successful and could develop their fluency with intentional practice. If this is true, then we must be intentional about providing a way for students to practice their fluency through repeated readings. The Six Minute Solution provides a way for students to do just this. This method was coined by Adams and Brown (2007) and is built on the foundation of peer-reading and feedback using the same text throughout the week.

The Six Minute Solution incorporates running records in a peer-to-peer context. Students are placed into purposeful pairings where they read together and track each other’s growth using a data chart to visually see when they are moving towards their individualized pre-set reading goals. This will take place each day and both the students and teacher will be able to monitor the reading fluency growth of all students instead of just the growth of the struggling readers who are progress monitored weekly as a part of the Response to Intervention^2 (RTI^2) which is used as multi-tiered system of identifying and supporting struggling students with more individualized small group instruction based on needs. This knowledge of weekly student progress will help show if a student begins to fall behind or plateau so intervention can take place before the universal screeners which occur in fall, winter, and spring. The goal is for all students, no matter their starting point, to demonstrate growth in their fluency rate as time progresses and the goal of the Six Minute Solution is to push all students forward – both in their progress and skill as readers, as well as in their social interpersonal communication skills as they rely on their partners during reading and are able to practice giving kind and constructive feedback.

Purpose of This Study

The purpose of this study is to examine the effectiveness of the Six Minute Solution program as a best instructional practice for improving the reading fluency and rate of all students – both those reading below grade-level standards and those who are already meeting grade-level expectations. This study reviews quantitative data that has been collected through the daily goal-trackers of each individual student, as well as progress monitoring data collected weekly by the teacher for the struggling readers in tier 2 of RTI^2, and monthly for those meeting grade-level expectations. This data was used to help determine whether the Six Minute Solution is a beneficial and evidence-based approach for classrooms in Middle Tennessee to build classroom-wide reading fluency no matter the starting point of the students.

Research Question

The question guiding this study was: 1. Does peer-mediated fluency instruction via the Six Minute Solution help to improve student reading fluency and rate? The study investigates the effectiveness of the Six Minute Solution on enhancing reading fluency and rate of all students in the third-grade classroom both those only in whole group classroom instruction, as well as students who receive additional supports and services.

Literature Review

Introduction

Third-grade literacy benchmarks highlight the need for effective approaches to improving the reading skills of every student at all levels. To do this, we must look to foundational reading skills such as fluency. If a student struggles to decode and make sense of written works, they do not have much cognitive load left to make meaning of what they read. Such is demonstrated in Scarborough's rope where each factor of reading is broken down into the braided strands of language comprehension, or word recognition – where reading fluency sits (Scarborough, et al., 2009). When fluency is lacking, students struggle to build their comprehension and there is noticeable cognitive effort having to be put into decoding (National Center on Improving Literacy, 2020). This shows up as a clear issue when we review third grade test scores considering the TN State Law regarding student retention (Brown & Mumpower 2023).

Students in third grade must build strong fluency throughout the year so that when it comes time for state testing, the students can perform well under the rigorous conditions demonstrating their abilities to read to learn. Additionally, looking beyond the end of year testing and third grade retention laws, there is research which shows that students who are not reading on grade-level by the end of third grade, are more likely to drop out of high school and, pushing this even further, high school drop-outs are more likely to be incarcerated than college graduates (Reading Partners, 2013). This demonstrates the importance of pushing students and building reading fluency skills to the best of our abilities before they leave the third grade.

Instructional Foundations of Reading Fluency

Researcher Rasinski (2024) focuses on helping struggling readers improve. He discusses how he began by teaching phonics and vocabulary to try to grow these students, but it did not prove to be enough. He then based his methods on research done by Samuels (1997) and attempted using repeated readings in his classroom. He stated that focusing on fluency through repeated reading helped his students to progress in their reading skills, even in previously unread texts, because their confidence was growing as readers (2024). He found that the use of repeated readings is a key component of literacy instruction that is often neglected. Samuels (1997) explained the foundational research for repeated readings and the focus on reading speed over reading accuracy to keep students from becoming so worried about making mistakes that they become overly cautious and their rate decreases. This shows, in part, the relation between reading fluency and automaticity as students can decode quickly.

This study defines repeated reading as Shanahan (2017) does, being that students read aloud short passages multiple times to reach a level of reading success. Shanahan (2017) discusses multiple ways to use repeated reading such as listening to audio tapes and then creating their own to match as proposed by Chomsky (1978) and working in fluid groups that change as reading success is met. The goal is for students to read the author's words accurately and to be able to read texts at about the same speed as natural speech so that reading sounds like language (Shanahan, 2017). Weakland (2021) adds to these models by including the fact that before students are released to practice their reading fluency with a peer, the teacher should model reading the passage fluently to provide an opportunity for students to hear how a skilled reader reads the passage so that they can emulate this themselves when they read. Not only is there freedom to structure repeated readings in various ways to fit specific classrooms, but the practice has also been shown to lead to better reading performance in students.

Reading Fluency and Student Outcomes

Reading fluency in early grades is widely seen as a predictor of later academic success because students who are able to read accurately and automatically are better able to engage with grade-level content. This is consistently demonstrated throughout research as we see fluency playing a foundational role in the development of comprehension thanks to the need for fewer cognitive resources being spent on decoding and instead being able to focus on making meaning (National Center on Improving Literacy, 2020; Rasinski, 2024; Samuels, 1997). It has often been seen that students who have stronger fluency skills are more likely to comprehend texts while their peers who struggle with fluency face more significant barriers to understanding the same texts (National Center on Improving Literacy, 2020; Wymer, 2022). This further extends to long-term educational indicators beyond the third grade.

Many states, Tennessee being one of these, rely on third grade reading proficiency as a benchmark for predicting the academic risk of the student in the future, all the way to graduation outcomes (Brown & Mumpower, 2023; Reading Partners, 2013). The standardized tests that are used to determine if third graders can progress to the fourth grade require that they read quickly, accurately, and with understanding. Students who struggle to read fluently will perform poorly on state testing even if they have the underlying knowledge they need to be successful (Brown & Mumpower, 2023; Rasinski, 2024). Due to the high level of importance of these state tests, several studies suggest that both below-grade-level readers and proficient readers would benefit from additional structured fluency practice (Cunningham, 2018; Marr et al., 2010; Wymer, 2022).

There are many research-based ideas on how to help students increase their reading fluency in the classroom before they encounter these high-stakes assessments at the end of third grade; one of these methods is the repeated reading strategy, as coined by Samuels (1997). The act of reading grade-level text passages repeatedly helps to build student fluency as well as confidence when they encounter further unknown text passages which they will be required to read. While many reading curriculums include an element of repeated readings, the goal is for students to be looking for answers or information. With the Six Minute Solution, the goal of these short, daily, repeated readings is simply to focus on improving reading fluency which will later lead to students being able to build greater comprehension.

The Six Minute Solution

The Six Minute Solution was created by Adams and Brown (2007) to provide leveled reading passages, partner-reading procedures, and progress monitoring tools so students can keep track of their goals. The goal of this program is to increase fluency and rate through daily repeated reading practice. Students read the passage for 1 minute and track how many words they read correctly in that time with the help of their partner. Students should be paired strategically to align students of specific instructional levels, so they are not so vastly different in skill that one partner receives no help and the other feels discouraged. Peer-coaching fluency intervention has been found to strengthen early literacy skills in students as they are able to work within their zone of proximal development (Marr, et al. 2010; Vygotsky 1978). Cunningham (2018) conducted a study with second grade students and found that the use of the Six Minute Solution led to overall increased reading fluency. Additional support for this program comes from Sopris West Educational Services (n.d.), which illustrates how the Six Minute Solution connects with federally funded literacy initiatives for striving readers directly targeting fluency instruction. The structure of the Six Minute Solution is built upon widely accepted best practices of teacher modeling, repeated readings, and practicing with a peer. Samuels (1997) provided foundational research for repeated reading increasing fluency and decoding accuracy which Shanahan (2017) further describes repeated reading as a consistently effect approach to improving student fluency skills. Weakland (2021) and Wymer (2022) state that repeated reading will help students to build greater comprehension of the text and understand what they have read more deeply. To help students realize this comprehension piece, the Six Minute Solution includes a retelling segment on the last day of reading each passage. The less cognitive energy which students must use to decode, the more cognitive power they have to create comprehension of a text and make meaning. By using repeated reading from the Six Minute Solution, students are given the opportunity to verbally practice reading each day with a peer and build upon their expression and decoding which leads to greater comprehension. In turn, each student is also able to listen to the text read aloud as they track their partner's reading each day providing yet another exposure to the text and opportunity to engage.

Conclusion

Research overwhelmingly shows that reading fluency is an important foundational reading skill that is directly linked to comprehension and academic success – both immediate and in the long-term, specifically in the third grade when students first transition from learning to read, to reading to learn. Research on repeated reading and fluency (Samuels, 1997; Shanahan, 2017; Weakland, 2021) shows that as students gain fluency skills, they free up the cognitive resources they need to understand the text they are reading. This, in turn, helps the students to perform better on state assessments – the results of which are tied to retention and future achievement (Brown & Mumpower, 2023; CMCSS, 2015; Reading Partners, 2013). Peer-coaching methods further assist in growth by providing structured practice within the students' instructional levels (Marr et al., 2010). The Six Minute Solution integrated each of these evidence-based practices with daily repeated readings, strategic pairings, and ongoing progress monitoring to help lead to measurable gains in reading fluency (Adams & Brown, 2007; Cunningham, 2018; Sopris West Educational Services, n.d.). There is a need for further large-scale research on the effectiveness of the Six Minute Solution, especially in large-scale settings though existing studies do suggest that this program proves to be an effective instructional model to improve reading fluency. This study built on that foundation by examining how the Six Minute Solution influences reading fluency and rate in a self-contained third-grade classroom in Middle Tennessee.

Methodology

Research Design

This is a classroom-based action research study to explore the effectiveness of the peer-mediated fluency instruction through the usage of the Six Minute Solution program. Quantitative data will be collected for all students using their individual data trackers as a part of the Six Minute Solution packet. This packet includes reading passages for each week, a data chart, and colored pencils for students to mark the mistakes their partners make each day. For the two students in RTI^2, data will also be collected each Wednesday using FastBridge CBMReading, a brief assessment which has students read aloud for one minute to calculate their WCPM and accuracy rate using a new passage the students have not previously read. This helps show authentic WCPM growth with non-repeated readings as affected by this reading intervention.

Setting

This study took place in a third-grade classroom in a public Title 1 Elementary School serving about 820 Kindergarten- 5th grade students in Middle Tennessee. All classes in this school are self-contained and serve between 15 and 30 students. The 3rd grade class used in this study contains 19 students who remain with the same teacher for the duration of the school day except for the two students receiving tiered RTI^2 services in another room and during special areas – P.E., music, art, and library – which rotates daily. Many students in this school come from either a military background or low-income families. This contributes to many students transferring in and out of the school throughout the year. The demographics of the student population is about 40% white, 30% black, 20% Hispanic, and 10% other background. There are 2-4 students in each classroom who are a part of the Collaborative Behavior Team or other classroom behavior plans which use a behavior checklist throughout the day and a check-in/check-out system to help keep these students from disrupting lessons with outbursts of undesirable or violent behavior. For this study, the Six Minute Solution was implemented as a part of the morning work immediately following morning announcements/breakfast and preceding further English Language Arts instruction.

Participants

There were a total of 19 students at the beginning of this study, however two students moved and their final data was unable to be collected. However, two students were transferred into this classroom later in the year who had also been using the Six Minute Solution since the beginning of the 25-26 school year so their data is also used in this study. Out of the 19 students whose data will be included, there are nine girls and ten boys. The expectation at the start of third grade is that students can read between 70-90 words per minute or they are considered at high risk. One student (Student A) began the year below this by reading only 39 words per minute. Students who are below the 40th percentile are considered at risk, and four students (Student A, Student B, Student C, Student D) were very close to this mark. Students who scored between the 40th and 70th percentile nationally are considered on grade level and ten students scored within this. The remaining four students were above the 70th percentile and thus exceeding grade-level expectations for the beginning of the year.

Student A, a Caucasian boy who began in the 10th percentile nationally and struggles with letter formation and phonetic spelling, receives Tier 3 RTI^2 services in an alternate 3rd grade classroom, and, Student B, an African American boy who has an IEP for speech due to a stutter, began in the 42nd percentile nationally and received Tier 2 RTI^2 services in the home classroom but was moved to Tier 3 in an alternate room after the winter screening. Student C does receive language services as she is considered an English Language Learner. All students receive the same Tier 1 whole group instruction throughout the day and small group instruction with the classroom teacher throughout the week.

Intervention

The duration of this action research study is duration of the 2025-2026 school year, with the comparison of beginning of the year, middle, and end of the year data from the FastBridge CBMReading screeners. On Mondays, the students whisper-read the entire passage before choral reading it with the teacher to have a model of fluency to follow. They will then discuss any unknown words and be provided with a definition and pronunciation. Each day the students will read with their pre-selected partner based on reading levels and record their WCPM on their tracker. On the final day (Friday) the students will retell what they remember from the passage after reading with their partner.

The goal is for the Six Minute Solution to take 6 minutes each morning, each partner has a minute to read followed by a minute to calculate their WCPM and mark their goal tracker, the other two minutes are to be used for preparing materials and then putting them away. This often takes closer to ten minutes as students find their partners, prepare materials, and put things away afterwards. Bi-weekly, the teacher meets with the students to discuss their reading goals to determine if they have been met and should set a new goal or if they should keep working towards this. They also complete progress monitoring with the teacher using FastBridge CBMReading either weekly (if in RTI^2) or monthly to gain another data point and help to determine an accurate WCPM. The materials that are needed each day are the individual student folders containing reading passages and goal trackers, colored pencils, and the timer.

Data Sources, Collection, and Analysis

The main source of data came from the fall, winter, and spring CBMReading universal screeners. Each individual student’s data tracking sheet was also considered. Due to the concern that students may not accurately record their data, progress monitoring data recorded by the teacher was also considered as a baseline and end of study assessment. For the students in RTI^2, progress monitoring occurred weekly, for other students it took place monthly. The starting, middle, and ending data points were compared to determine growth or regression of each student’s WCPM and trends for the class vs trends for the struggling readers to determine patterns and implement interventions.

Limitations

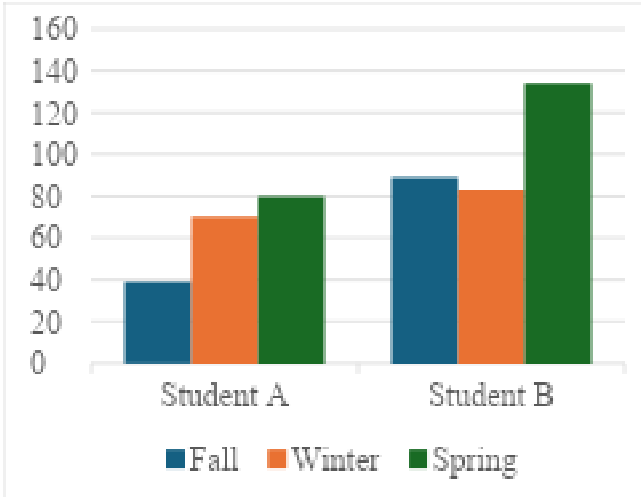
A limitation of this study is the class size giving a limited number of students' data to compare and identify patterns from. There were some inconsistencies in the pairings of the students as the groups had to be restructured due to personality conflicts, students moving schools or classrooms, as well as when students were absent or tardy, thus missing the Six Minute Solution and causing their partner to join a group of three. Absences also cause occasional lack of progress monitoring, leaving fewer data points to be considered. Finally, the accuracy and consistency of data recording by the students is questionable.

Findings

The goal of this study was to evaluate the effectiveness of peer-mediated fluency instruction through using the Six Minute Solution program in a self-contained third-grade classroom. To determine if this instructional model has proved to be effective in growing student fluency, the starting, middle, and ending word count of each student has been compared as determined through fall, winter, and spring universal screeners using FastBridge CBMReading. It was found that four of the 19 students experienced a decrease in WCPM from the fall to winter screener while six of the 19 students experienced a decrease from the winter to spring screener. Six students showed minor growth (10 words or less) between the fall and winter screener while the remaining nine students grew between 20 and 44 words per minute from the initial fall screen to the winter screener. From the winter to spring screener, five students showed minor growth and the remaining eight grew an additional 11 to 29 words per minute. Overall, from the start of the 25-26 school year to the end, three students (Student R, Student L, and Student Q) had a decrease in WCPM, two students (Student G and Student I) showed only minor growth, and the fourteen remaining students exhibited an overall growth between 14 and 45 words per minute. Eight of these students (Students A, B, C, E, F, J, M, and P) had an overall increase of 30 words or more per minute from the initial fall screener to the final spring screener.

Table 1

Students in RTI² Data from Fall to Spring Screener



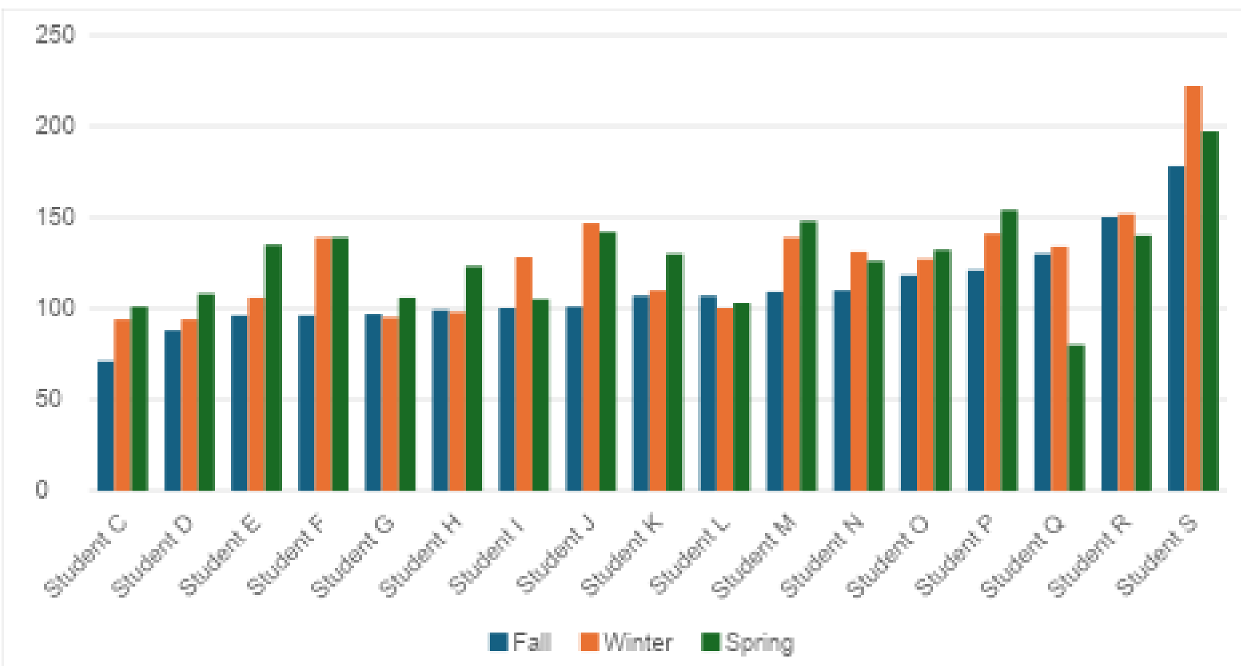
There were two students who received RTI² services during this study – Student A and Student B. Student A exhibited a growth increase of 41 WPM as they began with a WCPM of 39 had a WCPM of 70 at the winter screener and 80 WCPM in the spring. Another area of growth that has been seen for this student is that they are now more confident and volunteer to read or answer questions during whole group instruction, which was not the case before beginning the Six Minute Solution. Student B exhibited a growth increase of 45 WPM. At the fall screener, this student read 89 WPM, dropped to 83 WPM at the winter screener, but grew to 134 WPM by the spring screener. Due to the drop in WPM between fall and winter, Student B was moved from RTI² Tier 2 to Tier 3 which provided 15 additional minutes of instruction daily though, due to the growth displayed, has been moved back to Tier 2 RTI² services for the next school year.

Of the remaining students who receive no additional literacy support beyond Tier 1 whole group and small group instruction in their home classroom, three students have shown a decrease in WCPM from the fall to spring screeners. Student L dropped four words, Student R dropped ten words, and Student Q dropped 50 words. These students have been identified as needing additional targeted small group instruction to address this and will be entering RTI² or HDT next school year. The remaining 12 students showed some level of growth, from an increase between five and forty-five words each. Nine of these students gaining 20+ words per minute from the fall to spring screener as shown in table 2.

The students who displayed the greatest growth were Student A (39-70-80), who began with the lowest WCPM, Student B (89-83-134), Student F (96-139-139), and Student J (101-147-142).Through observing Students A, B, F, and J, it was seen that these students were very intrinsically motivated and were overjoyed each time they reached a new record. This helped to give them motivation and purpose to read quickly and accurately. This is also displayed in their desire to do their best on their classwork, asking clear questions and being engaged during instruction. Student F did reach a plateau as they grew unfocused and lacked motivation once state testing was over. The spring screener took place a week after state testing had concluded. This may have played a role in the scores of six students dropping from the winter to the spring screener.

Table 2

Data from Fall to Spring Screener for Tier 1 Students



The Six Minute Solution has helped all students in this third-grade class practice reading with the sole purpose of becoming more fluent, in setting, reaching, and reflecting on personal goals, and building their peer relationships and teamwork as they read and track together with a partner. Even though all students have not exhibited the same amount of growth in their WCPM, this classroom has seen an improvement in student enjoyment of reading and many students have begun building the intrinsic motivation of reaching their personal goals. The Six Minute Solution has built more conscious and willing readers for this specific classroom.

Many other classes and schools in Middle Tennessee have begun implementing the Six Minute Solution in various 3^{rd} – 5^{th} grade classrooms. In the school district where this study took place, there were about eight pilot schools implementing the Six Minute Solution at the beginning of the 2025-2026 school year. At the school where this study took place, the fifth-grade classes began implementing the Six Minute Solution upon returning from winter break with fourth grade being trained in the spring to follow suit at the beginning of the 2026-2027 school year with second grade possibly to follow. Seeing that the Six Minute Solution has provided growth in reading fluency for the students highlighted in this study, as well as an increase in enjoyment of reading in the classroom, more schools will follow suit with the implementation of this program throughout this Middle Tennessee school district.

Conclusion

The findings of this study established that the Six Minute Solution was an effective fluency instructional program to help most of the students in this third-grade classroom to increase their WCPM from the fall CBMReading screener to the spring CBMReading screener. Reading fluency is highly connected to reading comprehension and there has traditionally been a lack of opportunity for students to read for the sole purpose of building greater reading fluency, but the Six Minute Solution provided a resource to change this. As these third-grade students are able to develop greater fluency and rate, it builds upon their skills from second grade where they were focused on learning to read and connects this to the skills needed in the upper grade levels where they must now make the shift into reading to learn. The daily repeated readings help to build students' familiarity with the texts and increase their ability to read fluently as they would speak which, in turn, helps them to have greater cognitive load remaining to recall and create meaning from these texts.

The confidence these students gain and the fluency skills they develop proved to be a great asset to them when it came time for them to complete the high stakes end of the year state testing which, due to T.C.A. 49-6-3115, the scores they receive on the reading portion can cause them to be retained in the third grade. Even though there are interventions in place to help students who do not score proficiently on the state test, the implementation the Six Minute in third grade classrooms can provide the additional fluency support that students need before it comes to that point. Future research would be needed to investigate continued growth of students participating in the Six Minute Solution program. Additional research could also investigate if the percentage of students in classes who have been a part of the Six Minute Solution have to attend summer school or are retain is lower than the percentage of students in classes who have not used the Six Minute Solution. The shift of responsibility and increase in the workload from the second grade to the third grade, especially with the Tennessee retention law, is great so the support for students in their reading instruction should match.

References

Adams, G. N., Brown, S. M. (2007). The six-minute solution: A reading fluency program (primary level). *Sorpis West Educational Services*. ISBN 1-59318-591-X. Retrieved from https://people.wou.edu/~brownbr/Classes/The_Six_Minute_Solution/1_Six%20Minute%20Solution_PrimaryLvl/1_SixMinSolutnPrLvl_ppi-103.pdf

Brown, E., Mumpower, J. (2023). Student retention policies: Tennessee and other states (Legislative Research Analyst report*). Office of Research and Education Accountability, Comptroller of the Treasury*.

https://comptroller.tn.gov/content/dam/cot/orea/advanced-search/2023/Retentionfullreport.pdf

Chomsky, C. (1978). When you still can't read in third grade: After decoding, what? *In S. J. Samuels (Ed.), What research has to say about reading instruction* (pp. 13-30). *International Reading Association.*

CMCSS. (2015). 2014-2015 Mid-year review of strategic work. Retrieved from https://employees.cmcss.net/misc/ViewDocs?filename=annualReport_2014_MidYearReview.pdf

Cunningham, K. (2018). Exploring the effect of using the six-minute solution: A reading fluency program to increase overall reading fluency in second grade students. https://red.mnstate.edu/thesis/142/

Marr, M., Algozzine, B., Kavel, R.L., & Dugan, K. (2010). Implementing peer coaching fluency building to improve early literacy skills. *Reading Improvement,* 47(2), 74-91. Retrieved from https://files.eric.ed.gov/fulltext/EJ1317682.pdf

National Center on Improving Literacy. (2020). Why Does Fluent Reading Matter? https://www.improvingliteracy.org/resource/why-does-fluent-reading-matter

Rasinski, T. (2024). Reading fluency: The neglected key to reading success. *Education Week.* Retrieved from https://www.edweek.org/teaching-learning/opinion-reading-fluency-the-neglected-key-to-reading-success/2024/04

Reading Partners. (2013). Do prisons use third grade reading scores to predict the number of prison beds they'll need? Retrieved from https://readingpartners.org/blog/do-prisons-use-third-grade-reading-scores-to-predict-the-number-of-prison-beds-theyll-need/

Samuels, S. J. The method of repeated readings. (1997) *The Reading Teacher*, vol. 50, no. 5, 376-381. Retrieved from https://communityreading.org/documents/Samuels(1979)reprint_Repeated_Reading.pdf

Scarborough, H. S., Neuman, S., & Dickinson, D. (2009). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. *Approaching difficulties in literacy development: Assessment, pedagogy and programs*, 10, 23-38.

Shanahan, T. (2017). Everything you wanted to know about repeated reading. *Reading Rockets: Launching Young Readers.* Retrieved from https://www.readingrockets.org/blogs/shanahan-on-literacy/everything-you-wanted-know-about-repeated-reading

Sopris West Educational Services. (n.d.). Striving readers and the six-minutes solution: A reading fluency program alignment. *Cambium Learning Company.* Retrieved from https://www.kurzweiledu.com/files/Striving%20Readers%20and%20Six%20Minute%20Solution%20Alignment.pdf

The Nation's Report Card. (2022). NAEP Reading: Reading Results. *Nationsreportcard.gov.* https://www.nationsreportcard.gov/reports/reading/2024/g4_8/?grade=4

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. *Harvard University Press.*

Weakland, M. (2021). Extra! Extra! Repeated reading builds fluency. *Lesley University: Center for Literacy Teaching & Instructional Leadership.* Retrieved from https://centerliteracyteaching.lesley.edu/repeated-reading-builds-fluency/

Wymer, J. (2022). Using repeated readings to support fluency and comprehension. *Learning to Teach Language Arts, Mathematics, Science, and Social Studies Through Research and Practice,* 11(1). Retrieved from

https://openjournals.utoledo.edu/index.php/learningtoteach/article/view/547

Motivation, Self-Efficacy, and Academic Student Engagement in a First-Year Developmental Literacy Course

Amanda Bradley, Benita Bruster, Amber Kiser, Sloane Helmintoller, and Bethany Patterson

Austin Peay State University

Abstract

Many students enter the first year of college with their own personalized set of challenges, ranging from academic struggles, limited motivation, and a lack of self-efficacy. Students enrolled in developmental reading courses may experience challenges related to reading comprehension, academic writing, motivation, and self-efficacy. This study examines how students' perceptions of cognitive strategies, self-regulation, reflection, and confidence, and their anticipated future application of these skills, are influenced after participating in a college-level developmental reading course designed to support academic and social-emotional success. Grounded in Social Cognitive Theory (Bandura, 1997), this qualitative study used responses from 125 students via an open-ended survey to capture their perceptions of the course and its effectiveness upon completion. The results yielded an overall positive trend, with students reporting that although academic reading and writing within the course were challenging, the support of explicit and modeled strategies, along with instructional support, helped increase their understanding, motivation, and confidence to engage with the course material. Implications suggest that credit-bearing developmental literacy courses may benefit students who struggle with academic engagement, motivation, and confidence while fostering critical reading and writing skills necessary for post-secondary success.

Keywords: academic challenges, motivation, self-efficacy, first-year college students, developmental literacy

Introduction and Background

Students who are underprepared for college-level work may require greater motivation and persistence to successfully complete a degree (Bailey et al., 2010). For first-year college students, the transition to higher education presents significant academic and cognitive challenges, particularly for those enrolled in a developmental reading course. Students are expected to regularly engage in a variety of complex fiction and nonfiction texts, apply unfamiliar reading strategies, and complete writing tasks that demand analytical and critical thinking. Individuals who enter college underprepared for literacy demands often struggle to adapt to university coursework expectations (Chimenya, 2025). If prior gaps in reading comprehension and writing skills exist, the academic demands of college-level courses can lead to decreased motivation, lower engagement, and increased frustration (Bandura, 1997; Huang & Reynolds, 2022). As a result, understanding factors that influence student performance in developmental courses is critical for supporting academic success.

One factor that influences student success in academic settings is self-efficacy. Bandura (1997) defines self-efficacy as an individual's belief in their ability to successfully complete a task. These beliefs influence how students approach challenges, persist through difficulty, and engage in learning tasks. As explained by Bandura, perceived self-efficacy beliefs affect how consistently and effectively people apply what they know, which is a better predictor of intellectual performance than skills alone.

In addition to self-efficacy, motivation, and self-regulated learning, students' engagement with academic tasks and their persistence through difficulty are influenced by these behaviors. People develop beliefs about their capabilities, anticipate potential outcomes, and set goals based on those perceptions, which impact motivation. Self-efficacy plays a key role in regulating motivation (Bandura, 1997). The present study examines how academic challenges, motivation, and self-efficacy interact to influence student engagement and academic success in a first-year developmental literacy course.

Purpose and Research Question

Firmly grounded in Social Cognitive Theory (Bandura, 1997), this study encourages other higher education institutions to consider the benefits of a credit-bearing developmental reading course that models complex learning tasks, and the positive effects students derive from the course content. The purpose of this qualitative study was to analyze the benefits and challenges students reported from participating in a developmental reading course that focused on comprehensive reading and writing skills, scaffolded to foster confidence and motivation. The guiding research question was: How are students' perceptions of cognitive strategies, self-regulation, reflection, confidence, and their anticipated future application of skills influenced *after* participating in a college-level developmental reading course?

Literature Review

The transition to higher education presents academic, cognitive, and psychological challenges for first-year college students. Research suggests that these challenges, paired with student motivation, play a critical role in shaping students' willingness to engage with course material and apply effective learning strategies (Wigfield et al., 2016). Success in developmental literacy courses is determined not only by students' reading and writing abilities, but also by their ability to navigate academic challenges, sustain motivation, and develop confidence in their learning (Bandura, 1997; Schunk, 1994; Zimmerman, 2008). Students often encounter difficulties with increased academic rigor, unfamiliar expectations, and reading complex texts, all of which can affect student engagement and academic performance (Luther, 2022). At the same time, students' confidence in their academic abilities influences their persistence and long-term success, particularly when faced with challenging texts (Bandura, 1997; Zimmerman, 2008). Motivation also plays a role in students' engagement and academic achievement in literacy contexts (Luther, 2022; Schunk, 1994). The following sections examine the interconnected factors of academic challenges, student motivation, and confidence to provide a comprehensive understanding of how these elements influence students' engagement and achievement in developmental literacy contexts.

Academic Challenges for Underprepared University Freshman

Whether students are transitioning from high school to college or re-entering the educational setting several years post-high school, there are a variety of challenges and barriers that might impede their academic success. Understandably, the language style, complexity, and high-level vocabulary demands of college-level texts make comprehension difficult for students who might also have weak English skills (Anwar & Sailuddin, 2022; Stebleton & Soria, 2012). Due to this struggle with understanding academic texts, students also face the challenge of knowing how to approach writing tasks ranging from discussion board responses to multi-page essays (Insua, et al., 2018).

Additionally, possessing inadequate study skills, combined with limited reading practice and strategy usage when encountering complex reading and writing tasks, can cause students to struggle academically in all subject areas, which also impedes comprehension (Stebleton & Soria, 2012; Anwar & Sailuddin, 2022; Poole, 2014; Mulaudzi, 2023). Shifting from an external academic focus, some of these challenges can be internal on a social or emotional level, such as previous negative educational experiences or stress and anxiety related to the fear of academic failure (Stebleton & Soria, 2012; Mulaudzi, 2023). Even the large class sizes in a university setting, compared to the smaller class sizes of public or private schools, can be an adjustment for students. Additionally, many students are not accustomed to the lecture-based teaching style common in higher education (Anwar & Sailuddin, 2022; Mulaudzi, 2023). These academic-related challenges can accumulate and potentially lead students to question whether seeking an advanced degree is a goal they can truly achieve. Therefore, it is essential for educators to acknowledge the role motivation plays in student success and achievement.

Student Motivation

Reading motivation plays a critical role in students' ability to engage with academic texts, apply effective strategies, and persist through challenging reading tasks (Wigfield et al., 2016). Rather than functioning as a secondary outcome of instruction, motivation directly influences how students approach reading and whether they fully invest in the learning process. Their findings show that students must be willing to invest time and effort in developing reading skills and successfully applying comprehension strategies, highlighting motivation as a key factor in academic success.

Motivation is shaped by both students' goals and their level of interest in reading tasks. Intrinsic motivation, which reflects engagement driven by curiosity or personal value, is associated with deeper comprehension and long-term persistence (Wigfield, et al., 2016). In contrast, extrinsic motivation, such as completing tasks for grades or other external rewards, is often linked to more surface-level engagement and limited understanding. The difference between the two suggests that students are more likely to engage more meaningfully with academic texts when they see value in the task, rather than only completing assignments for external outcomes, such as grades.

Students' views of the value of their tasks and their beliefs about learning also influence motivation. Research conducted by Pintrich et al. (1991) discussed how the Motivated Strategies for Learning Questionnaire (MSLQ) identifies key motivational components, including intrinsic goal orientation, extrinsic goal orientation, and task value, which shape how students engage with academic tasks. The MSLQ continues to be widely used in higher education research to assess students' motivational beliefs and learning strategies. It focuses on constructs such as task value, goal orientation, and self-regulation, making it particularly relevant for understanding how college students engage with academic reading tasks (Pintrich et al., 1991).

Recent research with American college students further supports the importance of motivational constructs in reading. Kambara et al. (2025) examined reading motivation, reading amount, and reading comprehension among college students and found that motivational factors such as value and cost were connected to reading amount and comprehension. Their findings are especially relevant to developmental reading courses for first-year students because they extend beyond K-12 populations and show that students' motivation influences both how much they read and how they perform on reading-related outcomes.

Motivation is also closely connected to the self-regulation of learning processes. Schunk (1994) explained that effective self-regulation requires students to have goals, motivation, and self-efficacy for learning and performance. Self-regulation involves students' ability to set goals, monitor their understanding, and adjust strategies when needed (Zimmerman, 2008). These processes allow students to take an active role in their learning and persist when faced with challenges, reinforcing the connection between motivation and academic success.

Research has shown that metacognition is strongly connected to motivation and self-regulation in the development of reading proficiency. Early researchers such as Paris et al. (1984) investigated students' awareness of their cognitive abilities and learning strategies, while Baker and Brown (1984) examined the metacognitive processes involved in reading comprehension. Additional studies by Armbruster and Anderson (1982) and Garner (1987) further supported the idea that successful readers actively regulate their learning by monitoring comprehension, identifying important information, managing study time, and applying effective reading strategies. These findings suggest that literacy development requires more than decoding accuracy; students must also develop motivation, metacognitive awareness, and self-regulation skills that help them construct meaning from text and become independent readers (Israel et al., 2005). Overall, the research indicates that reading motivation is a key factor influencing student engagement, strategy use, and comprehension. When students are supported through clear instruction, structured strategies, and meaningful learning experiences, they are more likely to persist and actively engage with academic tasks. These findings provide a foundation for understanding how motivation operates within developmental literacy courses and directly inform the interpretation of the findings presented in this study. It also provides a look at how students develop greater confidence as they improve their reading skills and the role this plays in success in developmental courses and in overall academic success.

The Role of Confidence in Academic Environments

Bandura (1997) defines perceived self-efficacy as the belief in one's ability to successfully complete a task. Self-efficacy is not generalized to overall competence but reflects confidence toward a specific task, and individuals with low self-efficacy may avoid tasks. In contrast, those with higher confidence levels are more likely to engage in a particular activity (Artino, 2012). Research further suggests that students with higher efficacy beliefs exhibit better time management, greater persistence, and more accurate problem-solving, and that, in some contexts, confidence may predict performance better than skills alone (Bandura, 1997). These findings suggest that first-year college students with low self-efficacy may not put forth full effort academically when faced with the cognitive demands of college courses.

Supporting this, a systematic review by Honicke and Broadbent (2016) demonstrated a moderate relationship between academic self-efficacy and academic performance. Factors influencing this connection included effort regulation (persisting through difficulty), deep processing strategies (engaging meaningfully with content), and goal orientations (students' reasons for learning). These findings highlight the importance of self-regulated learning behaviors in supporting academic success. Although conducted in online environments, Broadbent and Poon (2015) identified strategies that may also benefit students in other postsecondary settings. Students who manage their time effectively, think critically about course topics, and persevere in understanding content despite challenges are more likely to achieve higher grades. In collegiate-level developmental reading courses, students may be required to read texts outside of class to better prepare for discussions, class content, and writing tasks, so students must be able to independently and actively engage with coursework. Students with stronger academic confidence may be better able to use these self-regulated skills more consistently.

From an instructional perspective, self-efficacy is malleable and can be developed. For students navigating the demands of a developmental reading course, intentional practices that strengthen student self-efficacy beliefs may help students persist through challenges and use more effective learning strategies, ultimately contributing to improved academic performance (Bandura, 1997).

To bridge the gap between academic challenges, motivation, and confidence, a postsecondary institution could utilize a targeted developmental literacy curriculum. Specifically, the reading course, RDG 1010, serves as an instructional framework that directly counters these academic, motivational, and self-efficacy barriers through intentional pedagogical scaffolding. By incorporating theoretical models of metacognition, self-regulated learning, and strategic reading into practical classroom experiences, the course curriculum systematically moves students from passive to active, scaffolds their learning, and ultimately develops independent learners. The following section outlines the instructional strategies, writing frameworks, and vocabulary models that form the core content and design of a credit-bearing, developmental reading course.

Content of Course

The course, RDG 1010, is a foundational college-level course designed to support students' reading comprehension, critical thinking, academic writing, and engagement with college-level texts across disciplines. Instruction included explicit modeling of reading strategies such as annotation, monitoring comprehension, and making connections, along with class discussions and weekly writing assignments. Students used the RACE structured writing format (Restate, Answer, Cite, Explain) to support evidence-based responses. Vocabulary development was also supported with the use of semantic maps and Frayer Models, which transitioned from guided practice to independent application throughout the semester. The regular implementation of these reading and writing strategies was intended to prepare students for the demands of postsecondary coursework.

Methodology

This qualitative study examined students' perceptions of a credit-bearing undergraduate developmental reading course. Data were collected at the conclusion of the Fall 2025 semester from students enrolled in RDG 1010 at a comprehensive state university in Middle Tennessee. The questionnaire included open-ended questions designed to capture student perceptions of the course.

Participants

Participants were first-year undergraduate students enrolled in a credit-bearing developmental literacy course at a public university in Tennessee. Developmental literacy courses in this context serve students identified as needing additional support in academic reading and writing based on institutional placement criteria. These courses provide explicit instruction in reading strategies and writing processes while fulfilling credit requirements. At the conclusion of the Fall 2025 semester, a survey was distributed to all the undergraduate students enrolled in RDG 1010. Of the 308 students enrolled, 125 students opted to participate and responded, yielding a 40.6% response rate.

Data Collection and Analysis

Students were provided a link to a questionnaire by the instructor and completed the questionnaire outside of class time. The electronic questionnaire included open-ended questions that asked students to identify challenges experienced in the course, perceived benefits, and future recommendations for RDG 1010. Participation was voluntary but highly encouraged. Writing samples were also collected from the same participant group; however, these samples are beyond the scope of the present study and will be reported separately. Responses were reviewed to identify recurring patterns across student feedback. Similar responses were grouped into thematic categories, and representative comments were selected to illustrate key findings.

Findings

As shown in Table 1, an analysis of student responses and survey results revealed several key patterns related to motivation, strategy use, and engagement in RDG 1010. To systematically identify, analyze, and report patterns in open-ended student texts, this study used a qualitative thematic analysis framework following the landmark six-phase approach established by Braun and Clarke (2006). Researchers analyzed data from repeated readings of student responses, and interpretations were used to identify themes based on patterns of the data. Themes were reviewed and refined to ensure an accurate depiction of the results and then student responses were chosen to articulate student perceptions and experiences while enrolled in the developmental reading course. Descriptive feedback indicated that students’ perceptions of cognitive strategy use, self-regulation, reflection, confidence, and anticipated future application of skills were relatively high. This connection is especially important because the course is designed to teach strategies and to help students transfer strategies to other coursework.

The RDG 1010 course findings demonstrate how student motivation manifests in a real-world classroom setting. Students reported challenges related to complex texts and writing demands, which often created feelings of overwhelm and reduced engagement. At the same time, students identified instructional support, reading strategies, and interactive learning opportunities as key factors that increased their motivation and engagement. These findings suggest that motivation is an internal characteristic shaped by instructional design and the environment.

The intersection of these findings highlights a critical link between instructional intervention and student emotional states. When confronted with academic obstacles without proper scaffolding, students quickly experienced cognitive overload, which directly degraded their motivation. However, when instructors introduced interactive learning opportunities and explicit strategy modeling, students successfully converted that anxiety into academic confidence. This shift underscores that strategic intervention in developmental literacy courses serves as a buffer against academic frustration, effectively bridging the gap between students' perceived task difficulty and their intention to persist in their academic career.

Table 1

Qualitative Themes from Student Responses

Theme	**Definition**	**Representative Examples**
Challenges in Writing and Text Comprehension	Students described difficulties with writing assignments and understanding course readings. Responses reflected challenges in managing the cognitive demands of reading comprehension and written expression simultaneously,	“Writing assignments were difficult.” “Some readings were hard to understand.”

	which may have influenced engagement and academic performance.	"I struggled to put my thoughts into writing."
Helpfulness of Reading Strategies and Writing Supports	Students identified structured reading strategies and writing supports as beneficial for improving comprehension and assignment completion. Responses suggested that tools such as annotation, guided frameworks, and organizational supports increased confidence and engagement.	"Annotation helped me understand the reading better." "The writing frameworks made assignments easier." "Breaking the readings into steps helped me stay focused."
Instructional Approach and Student Self-Efficacy	Students emphasized the importance of instructional support, interactive teaching methods, and classroom environment to promote engagement and confidence. Responses suggested that supportive instruction contributed to increased self-efficacy and success.	"The instructor explained things clearly." "Class discussions helped me understand the material." "I became more confident in my reading and writing abilities in the course."

Note. Themes were derived from student responses collected in RDG 1010.

Discussion

These findings contribute to the understanding that student motivation and confidence play a key role in the success of first-year students enrolled in a developmental reading course. Instructors of these courses should incorporate instructional strategies that model and promote critical thinking, scaffold complex reading and writing tasks, and encourage engagement with academic texts. With this support, students may be better equipped to navigate academic challenges in higher education while building confidence. Additionally, the qualitative insights gathered from RDG 1010 reinforce the core components of Social Cognitive Theory (Bandura, 1997) by showing that when instructors intentionally implement explicit modeling, scaffolded reading and writing tasks, and interactive critical thinking strategies, they directly help alleviate cognitive overload often triggered by challenging postsecondary texts. This targeted pedagogical support provides the essential framework for transitioning underprepared first-year students from passive consumers of information to self-regulated learners. Ultimately, these results align with and extend existing research, suggesting that when institutions look past basic skill deficits to holistically address a student's academic confidence, first-year students are substantially better equipped to overcome transitional barriers and persist through the rigorous demands of higher education.

Limitations

The findings of this study should be interpreted with consideration of a few limitations. First, there was a low percentage of student participation in the research compared to the number of students enrolled. Only 40.6% of students volunteered to participate. With a larger sample size, there would have been a broader understanding of the helpfulness and challenges presented in the course. Additionally, data collection only occurred at the end of the semester. In two different areas (Challenges and Suggestions for Future RDG 1010 Courses), many participants reported experiencing no challenges and offered no suggestions for future course offerings. This could be interpreted as positive, indicating the course was taught at a level that made it accessible to students, or as negative, indicating they did not feel they grew due to a lack of challenge. Additionally, the lack of suggestions for future RDG 1010 courses could indicate that the course was satisfactory and met expectations, or that students were unsure how to advocate for and voice their opinions or concerns.

Implications

Based on the positive results of the qualitative study, other universities could benefit by offering a credit-bearing developmental course focusing on reading comprehension, academic writing support, and increasing students' motivation and confidence. Building a strong foundation in these areas can support students as they navigate the challenges of completing a degree program. Findings suggest that when students are explicitly taught metacognitive reading strategies, their awareness of their use increases (Iwai, 2016). Our study supports these findings, as participants were taught these strategies and skills and reported benefits; therefore, integrating a program like this into programming for underprepared students at the community college or university levels would be ideal.

Conclusion

Ultimately, students entering college continue to benefit from increased motivation and confidence when encountering a variety of academic challenges. As Bandura (1997, p. 77) famously quotes, “Self-belief does not necessarily ensure success, but self-disbelief assuredly spawns failure.” Many students enter college with extensive self-doubt, and if allowed to fester, it will ultimately lead to failure; however, educators hold the keys to fostering self-belief in students by teaching them ways to navigate writing and text comprehension using supportive strategies to develop student self-efficacy. The decision to use these strategies independently lies within students’ control, but fostering self-efficacy increases the likelihood of achieving the paramount goal of graduation. University-level education differs from the free, appropriate, public education provided in elementary school through high school; nevertheless, underprepared college students must be supported in reaching their fullest potential. By offering a developmental reading course focusing on teaching comprehension and critical thinking skills that support student success in reading and writing, students will specifically improve their ability to analyze text, understand academic material, and communicate ideas effectively. These outcomes build confidence, motivating students to achieve their academic goals as they embark upon their journey in higher education.

References

Anwar, I. W., & Sailuddin, S. P. (2022). Academic reading difficulties in higher education. *JOLLT Journal of Languages and Language Teaching*, *10*(2), 309–314. https://doi.org/10.33394/jollt.v10i2.4849

Armbruster, B. B., & Anderson, T. H. (1982). *Idea-mapping: The technique and its use in the classroom or simulating the "ups" and "downs" of reading comprehension* (Reading Education Report No. 36; ED221842). ERIC. https://eric.ed.gov/?id=ED221842

Artino, A. R. (2012). Academic self-efficacy: From educational theory to instructional practice. *Perspectives on Medical Education*, *1*(2), 76–85. https://doi.org/10.1007/s40037-012-0012-5

Bailey, T., Jeong, D. W., & Cho, S.-W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, *29*(2), 255–270. https://doi.org/10.1016/j.econedurev.2009.09.002

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 353–394). Lawrence Erlbaum Associates.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Chimenya, U. M. (2025). Transition from high school to university: First-year students' reading experiences. *Reading & Writing*, *16*(1), 1–10. https://doi.org/10.4102/rw.v16i1.517

Garner, R. (1987). Strategies for Reading and Studying Expository Text. *Educational Psychologist*, *22*(3–4), 299–312. https://doi.org/10.1080/00461520.1987.9653054

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review*, *17*, 63–84. https://doi.org/10.1016/j.edurev.2015.11.002

Huang, S., & Reynolds, M. (2022). Factors that influence college students' reading motivation. *Athens Journal of Education*, *9*(2), 187–210. https://doi.org/10.30958/aje.9-2-1

Insua, G. M., Lantz, C., & Armstrong, A. (2018). Navigating roadblocks: First-year writing challenges through the lens of the ACRL framework. *Communications in Information Literacy*, *12*(2), 86–106. https://doi.org/10.15760/comminfolit.2018.12.2.3

Israel, S. E., Block, C. C., Bauserman, K. L., & Kinnucan-Welsch, K. (Eds.). (2005). *Metacognition in literacy learning: Theory, assessment, instruction, and professional development*. Taylor & Francis.

Iwai, Y. (2016). The effect of explicit instruction on strategic reading in a literacy methods course. *International Journal of Teaching and Learning in Higher Education*, *28*(1), 110–118. https://files.eric.ed.gov/fulltext/EJ1106323.pdf

Kambara, H., Lin, Y.-C., Lin, H.-C., & Chen, P.-Y. (2025). Examining the mediating role of reading amount between reading motivation and reading comprehension among American college students: Ethnic similarities and differences. *Social Psychology of Education*, *28*(1), 1–33. https://doi.org/10.1007/s11218-025-10034-8

Luther, V. L. (2022). The impacts of self-efficacy and intrinsic motivation on reading achievement. *Journal of Inquiry and Action in Education*, *13*(1), 54–71.

Mulaudzi, I. C. (2023). Challenges faced by first-year university students: Navigating the transition to higher education. *Journal of Education and Human Development*, *12*(2), 79–87. https://doi.org/10.15640/jehd.v12n2a8

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology*, *76*(6), 1239–1252. https://doi.org/10.1037/0022-0663.76.6.1239

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)* (Technical Report No. NCRIPTAL-91-B-004). University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning. https://files.eric.ed.gov/fulltext/ED338122.pdf

Poole, A. (2014). The strategy use of struggling readers in the first-year composition classroom: What we know and how we can help them. *The CEA Forum*, *43*(1), 3–32. https://ceaforum-ojs-ttu.tdl.org/ceaforum/article/view/7037

Schunk, D. H. (1994, April 4–8). *Student motivation for literacy learning: The role of self-regulatory processes* [Paper presentation]. Annual Meeting of the American Educational Research Association, New Orleans, LA, United States. https://eric.ed.gov/?id=ED367676

Stebleton, M. J., & Soria, K. M. (2012). Breaking down barriers: Academic obstacles of first-generation students at research universities. *The Learning Assistance Review*, *17*(2), 7–19.

Wigfield, A., Gladstone, J. R., & Turci, L. (2016). Beyond cognition: Reading motivation and reading comprehension. *Child Development Perspectives*, *10*(3), 190–195. https://doi.org/10.1111/cdep.12184

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and prospects. *American Educational Research Journal*, *45*(1), 166–183. https://doi.org/10.3102/0002831207312909

Perceptions from the Field: Decision-Making Curriculum in the Secondary Social Studies Classroom

Jeffrey M. Byford, Ph.D.
The University of Memphis

Alisha Milam, Ed.D.
Mississippi State University - Meridian

Abstract

This qualitative case study examines high school social studies teachers' perceptions of and use of Robert Stahl's values-based decision-making materials in standards-driven classrooms. Drawing on semi-structured interviews with seven teachers from diverse regions, the study explores how decision-making curricula support inquiry, historical empathy, and student engagement. Findings reveal three central themes: (1) teachers value decision-making as a meaningful pedagogical approach that promotes higher-order thinking and student-centered learning; (2) standards-based accountability and high-stakes testing constrain teachers' ability to implement inquiry-based practices; and (3) social studies is often deprioritized relative to tested subjects, limiting access to instructional resources. Despite these barriers, participants viewed Stahl's framework as an accessible and effective way to integrate inquiry into existing curricula to better prepare students for civic engagement and critical thinking.

Keywords: Decision-Making, Values-Education, and Historical Empathy

Introduction

I believe that our failure to focus on decision-making in our history classrooms reflects the state of society today. We have students who look to others for answers and struggle to make decisions independently. Incorporating decision-making activities and lessons about historical events makes learning more engaging because students should be actively participating rather than just taking notes or completing worksheets.

When asked to explain the importance of including decision-making materials in her classroom, Laura, a United States history teacher with 13 years of experience, expressed dismay and frustration that more teachers do not incorporate value-based decision-making materials into their curricula.

Laura and many other history teachers repeatedly expressed a desire to use cognitive-based learning techniques to engage students in history. Historical periods are often presented through minimal lectures, unremarkable readings, and passive worksheets. This approach yields little student interaction or in-depth analysis, limiting opportunities for higher-level learning (Rapopart, 2006; White, 1994). Overall, students value lessons that go beyond direct instruction and rote memorization; however, year after year, student understanding of United States and World History remains vague and confusing.

This conundrum becomes especially clear when teaching concepts that seem irrelevant to students' lives. Participants indicated a need for more student-centered activities that promote inquiry-based decision-making and foster historical empathy. This goal of providing decision-making opportunities for all learners aligns with Robert Stahl's four-phase model of decision-making (Stahl, 1995). Stahl emphasized that cognitive psychology is key to effective teaching and learning, supporting the idea that no one is born with prior knowledge and understanding; instead, knowledge is acquired by observing the environment.

Previous research shows that while history teachers value cognitive-based learning techniques, many feel pressured to modify curriculum and instruction to promote student success and equitable access to education, often linked to accountability measures such as test scores (Popham, 2006). At the national level, the U.S. Department of Education (DOE), following guidelines from No Child Left Behind (NCLB), regulated the historical topics and themes within content areas. Although NCLB did not specify pedagogical methods for meeting its standards, it emphasized high-stakes testing, which put pressure on educators nationwide to focus on tested material in their teaching and curricula. This has led many school systems in the U.S. to dedicate more time to heavily tested subjects, resulting in the replacement of cognitive learning with a mandatory focus on testing skills and content (Pratt & Richard, 2019). One result is that teachers often attend seminars and professional development programs that are less supportive of educators and offer little clear benefit to students. Additionally, many states and districts prioritize standardized testing to determine a school system’s future operations and funding, rather than emphasizing the development of knowledge or decision-making skills for future citizens.

Despite increased restrictions on pedagogical strategies related to teaching to the test, participants in this study identified Robert Stahl's decision-making materials as an effective way to incorporate inquiry, value clarification, and historical empathy into the high-stakes testing environment. For example, teachers prefer that students self-identify decision-making phases as a natural and essential part of knowledge construction. Teachers also want students to relate to scenarios and content materials during decision-making. These tendencies align with Thomas Conklin's and Artemis Boulamatsi's (2020) research, which suggests that decision-making exercises can be used across various disciplines and for content retention by creating, evaluating, analyzing, and responding.

Building on prior work implementing Robert Stahl's content-centered learning strategies for historical, values-based dilemmas and decision-making (Agarwal & Roediger, 2019; Sheurman, 2018), this study presents an example of how the decision-making aspects of content-centered learning can serve as an additional curricular option to enhance values clarification and critical inquiry through a series of lessons centered on historical events and dilemmas. The effect of these lessons on history teachers' ability to contextualize historical events is examined through a qualitative case study approach.

This study aims to answer the following question: What are high school teachers' perceptions of and their use of Robert Stahl's decision-making materials when teaching history? Because these materials are designed to promote inquiry, clarify values, and encourage cooperative learning, we expected that the interviews would lead to greater use of supplemental resources to support curriculum development and content retention.

Theoretical Framework: Decision-Making with Content-Centered Learning

Teaching about values and incorporating inquiry skills to promote student learning of complex or controversial ideas in the social studies curriculum is not a new phenomenon among teachers. Most social studies teachers use instructional strategies centered on school district pacing guides and sample test materials to guide their pedagogy. However, because no priority has been placed on inquiry and project-based learning activities across Tennessee, a resurgence of Stahl's influence and pedagogical imprint may be influential. Stahl's values-based decision-making education model adapts to any social studies content and allows teachers to adhere to state standards while concurrently employing inquiry and project-based learning activities.

Historical Empathy and Decision-Making

American high school teachers have faced challenges in adopting innovative pedagogical strategies to deepen understanding of historical context (Berman, 1986; Dunn, 1983; Schlene, 1996; Pham & Hayden, 2019). Previous research has sought to measure students' attitudes toward the curriculum and socio-political issues to foster more critically aware students who can make decisions and analyze value dilemmas, including hypothetical scenarios (Ahmad, 2017; Balakrishnan, 2009). Recent curriculum reforms, primarily through the gradual integration of supplemental lessons and activities focused on decision-making and values, may improve students' ability to discuss topics such as war, political events, and causes that some might find controversial (Lennon, Byford, and Cox, 2015; Cowgill and Waring, 2017).

The term historical empathy, or perspective-taking, is often defined as 'the ability to see and entertain as conditionally appropriate connections between intentions, circumstances, and actions, and see how any particular perspective would have affected actions in specific circumstances' (Lee & Ashby, 2001, p. 25). A common misconception about historical empathy is that it depends on the student's cognitive skill level. Boddington (1980) suggests that empathy is a complex mix of thinking and feeling. Empathy cannot be achieved without the student's positive attitude and the cognitive skills to understand other potentially significant points of view.

Portal (1987) suggests that historical empathy is the foundation of historical inquiry. To practice historical empathy, a student must be able to: 1) project their thoughts and feelings into a historical situation or moral dilemma; 2) distinguish the historical period from their own; 3) use references and materials to support the dilemmas, when applicable; 4) present the person(s) or situation to illustrate the circumstances of the case or moral dilemma; and 5) be aware of misunderstandings, conflicts, or tragedies.

Recent trends in social studies curriculum design, highlighted by a resurgence in values clarification and inquiry, promote student discussion of historical events and controversial issues (Cowgill & Waring, 2017; Lennon, Byford & Cox, 2015). Along with a greater emphasis on values education, each standard within the current NCSS (2019) C3 framework challenges a linear approach to teaching social studies and recognizes the importance of skills, knowledge, and investigations as a foundation for social studies content. The C3 framework promotes an inquiry-driven classroom where students engage in civic life and incorporate thought-provoking discovery into social studies.

Robert Stahl's Four Phases of Decision-Making

Robert Stahl argued that cognition depends on a strong foundation of background knowledge and information. He advocated integrating a decision-based, content-focused curriculum into any existing instructional plan to promote classroom growth. According to Stahl (1976), learning and knowledge retention require higher-order thinking supported by decision-making materials. Therefore, a content-focused curriculum developed by J. Doyle Casteel and Robert Stahl enables classroom teachers to "carefully plan for, develop, implement, monitor, and assess instructional activities" (p. 7).

A key feature of the content-centered curriculum was the integration of cognitive psychology. Robert Stahl (1995) argued that "human knowledge, whether personal or associated with a field or discipline, is constructed internally within and by each human" (p.7); this foundation of learning was rooted in cognitive psychology. Stahl maintained that authentic learning results from the construction of knowledge. The term construction refers to a process of building that supports the "perspective claim that individuals spend their lives constructing or inventing their meanings, ideas, views, and conceptions of the actual and imagined worlds, and interpretations of the phenomena they encounter" (p.14). Consequently, students interpret, analyze, and enact their values, enabling them to take ownership of the decision-making process.

Casteel and Stahl (1997) argued that students engaged in values-based activities progress through four distinct decision-making stages as they explore and make key choices. Content-specific activities aim to develop students' use of conceptual, relational, moral reasoning, and reflective thinking. Asking students to analyze, evaluate, and draw values-based conclusions is not new to social studies education. Moral and value dilemmas date back to the 1960s, with the Harvard Social Studies Project, which sought to involve students not merely as passive spectators but as active thinkers in history and contemporary life. The Harvard Project did not provide predetermined correct or incorrect answers to social and historical issues. Instead, it challenged students to form their own positions and navigate conflicting societal viewpoints (Levin, Newmann, & Oliver, 1969). Like other projects from the New Social Studies era, the Harvard Project eventually lost prominence; however, some elements of its materials persisted in isolated classrooms.

The initial thinking stage occurs at the conceptual level. This phase focuses on student understanding as they answer explanatory questions such as who, where, when, what, how much, and how many. When moral issues or problems are explored and decisions are made, moral reasoning is emphasized. Students are expected to use descriptive language patterns during this phase. These patterns demonstrate their understanding of the situation, problem, or dilemma. Therefore, they identify the nature of the issue or dilemma, including specific moral concerns and elements. In line with Stahl (1979), students engage in conceptual thinking by answering questions such as: What does this mean? What is your interpretation of the situation or events? What data is known? Also, what are some relevant attributes, characteristics, or features discussed? These questions aim to assist with the initial steps of identifying potential objects, situations, and problems under examination.

The second stage of thinking and decision-making occurs at the relational level. This phase emphasizes methods and approaches; the teacher may help students engage in moral reasoning within the subject's context and content. The relational phase highlights the application and analysis of the activity and course materials. Analyzing and applying data and materials encourages the integration of these elements, leading to the combination of learned material with new meanings and concepts (Casteel & Stahl, 1997; Stahl, 1995). During this phase, students aim to connect their understanding of morality with the ideas, content, images, and perspectives related to the issue.

The moral reasoning stage, or the third phase, occurs when students engage in moral reasoning. Students use moral criteria to evaluate and decide which consequences they want to achieve or protect. These criteria include the measures used, how requirements are applied, which policies are followed, which situations are considered moral, and which courses of action are taken. Students apply ethical standards when making decisions or judgments during this phase. They consider the consequences of their moral choices and whether those effects are justified. During this stage, students might improve their decision-making by asking questions such as, "What course of action would I take in this dilemma?" or "What results should I expect if I choose the first or second option?"

The final reflective phase occurs after the other stages are considered complete. This phase aims to help students evaluate how consistently they applied moral criteria and how they reached their moral judgments, usually through a public statement of their final decision and the reasons behind it. According to Casteel and Stahl (1997), "under optimum conditions, this phase of thinking also embraces [d] careful reconsideration and reprocessing of one's ideas, information, and beliefs" (p.18).

Content-Centered Learning

Predicting the four stages of decision-making is essential to content-centered learning. A foundation of content-centered knowledge supports value clarification and moral growth. Specifically, value clarification involves students' statements that demonstrate their engagement in understanding, conceptualizing, and clarifying values (Stahl, 1979). To promote both value clarification and moral development, content-centered learning typically includes one of these types of dilemmas: 1) a neutral issue or context that a student might respond to or consider from a moral perspective (e.g., building a local prison); 2) a moral or value issue presented in its extreme forms (e.g., someone being either lawful or unlawful, or honest or dishonest); 3) a moral or value conflict in decision-making (e.g., fighting for racial justice or choosing whether to spend money on fighting climate change while exploring for oil); 4) a moral or value issue that sparks debate because it involves two or more possible choices (e.g., a person who supports freedom of speech deciding what to do during a peaceful march of the American Nazi Party); and 5) a situation where multiple values or moral dilemmas may conflict (e.g., divorcing a spouse to marry someone else).

Stahl (1978) argues that content-centered activities stem from a social studies event or situation and align with one or more of the five types of dilemmas. Once an event, historical dilemma, or controversial element is identified, the teacher can select from five distinct decision-making formats for students to use. These strategies, developed by Robert Stahl (1995), help students resolve conflicts and determine appropriate actions or responses. Ultimately, the rank-order, forced-choice, negotiation, invention, and exploration strategies provide higher-order thinking opportunities for students and unique opportunities to promote critical pedagogy tailored to the context, conditions, and specific needs of the situation.

Method

In Tennessee, United States history is the only course tested at the high school level. United States history teachers have struggled to find innovative pedagogical strategies to help students understand historical context (Berman, 1986; Dunn, 1983; Pham & Hayden, 2019; Schlene, 1996). While previous research has attempted to measure students' perceptions and attitudes regarding curriculum and avenues for encouraging critical thinking, as well as to analyze value dilemmas, including hypothetical situations, this study investigates the impact and potential uses of supplemental decision-making and values-laden activities (Ahmad, 2017; Balakrishnan, 2009; Stahl, 1976, 1979).

Research Design

Five lessons on historical topics were developed using features of Stahl's decision-making, content-centered curriculum. To assess their potential effectiveness and usefulness as supplementary resources, a qualitative case study approach was used to gather participant insights into teaching diverse groups, experiences with the educational system, and opinions on the value of Robert Stahl's curriculum. This design enabled a thorough examination of the phenomenon in its context from participants' perspectives (Merriam & Grenier, 2019; Stake, 2017; Yin, 2017). Six semi-structured interview questions were developed to elicit participants' perceptions related to four research questions. Each question included two sub-questions aligned with the main questions, providing opportunities for participants to share their views.

Participants

The Tennessee Department of Education divided state school districts into three regions based on geographic, political, and cultural boundaries. Twenty schools were identified in the Eastern region, thirty-five in the Middle region, and twenty in the Western region. For sampling, each school district was treated as a stratum, and each district's secondary school(s) as a sub-stratum. The first stage selected 15 districts at random in each region. The second stage involved randomly selecting seven participants from the total of 45 districts using Excel's random function. These seven teacher samples, selected through a two-stage random sampling process, reflected the diversity of schools and teacher characteristics (see Table 1). Participating teachers attended a one-hour online training session designed to familiarize them with the four decision-making stages and to provide a brief overview of the four decision-making formats. These sessions were conducted by the researchers. Because the researchers relied on teachers' willingness to teach content-centered lessons aligned with a state-mandated curriculum, participants were provided with five history lessons developed for the study. These lessons cover various historical events, from progressivism to Social Darwinism.

Table 1.
Teachers' Characteristics

Name (pseudonym)	Gender	Age	Work Experience	Course	State Region
Rebecca	Female	47	25	World History	Eastern
Ken	Male	71	17	World History	Eastern
R.J.	Female	35	14	U.S History	Middle
Lance	Male	46	12	U.S. History	Western
Laura	Female	37	13	U.S. History	Central
Austin	Male	29	5	U.S. History	Central
Sarah	Female	33	7	World History	Middle

The Lessons

The first lesson covers the end of World War I. In the activity “14 Points,” students take on the roles of U.S. Senators. They examine the League of Nations and the Treaty of Versailles, focusing on a key element of President Wilson's Fourteen Points. Students analyze this element and decide on a logical outcome based on the needs of the U.S. and its allies. The second lesson, "Giving Them a Voice," has students act as members of a large automobile company's board of directors. They must decide how much influence workers should have over wages, working conditions, and production. Each student is given nine options for worker participation in company decisions and must rank them from best to worst, providing reasons for their choices. The third lesson explores the teaching of evolution through the question, “What is All this Monkey Business about Anyways.” Here, students investigate the 1920s Scopes Monkey Trial in Tennessee. The fourth lesson highlights the progressive movement and societal reform in the early 1900s. In "On Being an Important Person," students explore aspects of progressive activism and identify policies and traits that benefit American society. The fifth lesson, ‘Out of School”, looks at U.S. foreign policy after the 1905 Russo-Japanese War. As school board members, students vote on controversial policies, including the segregation of Asian and Caucasian children in San Francisco's public schools.

Table 2.
Robert Stahl's Decision-Making

PHASE	EXPLANATION
CONCEPTUAL	The *conceptual phase* in the decision-making process focuses on student comprehension of the task at hand. Explanatory answers to questions of who? where? when? what? How much? And how many apply? Conceptual thinking

	focuses on the comprehension, description, and clarification of information provided to the student
RELATIONAL	The *relational phase* stresses the formation, relevance, and connection to content materials used in the classroom. The relationship between content material is combined through a lesson or activity.
DECISIONAL	The *decisional phase* involves a student using categories of information as they consider all options before making a decision. Categories include given, likely, and possible choices, actions, and policies, as well as the probable short- and long-term advantages and disadvantages of each option or plan.
REFLECTIVE	The final *reflective phase* occurs when the other stages are presumed complete. In the reflective aspect, a student must include a public statement of a final decision and the rationales used to arrive at that decision.

Note: Adapted from Doorways to Decision Making by Casteel, J., & Stahl, R. (1997). *Doorways to Decision Making*. Prufrock PR.

Data Collection and Analysis

To assess the reliability and integrity of the research, all teachers reviewed the five lesson plans and a summary of Robert Stahl's four decision-making phases (see Table 2). They also indicated whether all lesson activities were completed successfully and whether any irregularities occurred. Interviews were conducted over a two-month period, from April to May 2025. The interviews took place at convenient times for each of the seven participants, all after school hours. The researchers used a semi-structured interview format for the recorded sessions. The questions (see Table 3) were based on the four questioning categories identified by Strauss et al. (1981).

Document analysis of participant responses was first conducted independently by one of the authors, who holds a doctorate in social studies education and has nine years of experience teaching United States and World History. Subsequently, the second author, also holding a doctorate in social studies education and with over 20 years of teaching experience in high school and university settings, reviewed the analysis. All discrepancies in responses were discussed until consensus was reached on the final document analysis, which identified themes. Themes were derived from individual cases as well as through a comparative synthesis. The reliability of the data was confirmed through triangulation, including field notes and document analysis.

Table 3.

Participant Interview Questions

1. What do you think is the ideal approach to presenting content to students?
2. Some might say that Robert Stahl's four decision-making phases are unnecessary for promoting student inquiry. What would you say to them?
3. Suppose your administration asked you to teach content that dealt with decision-making. What would it be like?
4. Some might say Robert Stahl's approach to teaching decision-making in social studies is outdated. What would you say to them?
5. How do you think teachers should address decision-making in the classroom?
6. Suppose your district required you to explain Robert Stahl's decision-making material as they relate to the state standards. What would you say to them?

Results

Three themes emerged from data provided by high school social studies teachers. Participants indicated that: a) decision-making materials developed by Robert Stahl are valued as an effective teaching method; b) standards-based curriculum and accountability act as barriers to student inquiry; and c) teachers perceive social studies education as deprioritized.

Theme 1: Decision-Making Materials Are Important

All seven participants interviewed expressed support for inquiry-based learning. Five of the seven teachers explicitly identified Robert Stahl's decision-making materials as an effective way to incorporate inquiry into the social studies classroom. When asked about their ideal approach to teaching social studies content, teachers highlighted multiple inquiry applications, all of which align with decision-making as the most desirable method. Teachers prefer to recognize decision-making phases as a natural and essential part of knowledge building. Participants consistently expressed a desire for students to have more student-centered opportunities in the classroom to engage in inquiry-based decision-making. This desire to create opportunities for decision-making for all learners aligns with Stahl's (1995) foundational concept of the four phases of decision-making.

From different regions of the state, R.J. and Lance identified Stahl's four decision-making phases as essential for learning. R.J. immediately recognized these phases in her students once she was introduced to the concept. "I think they are just based on what I see with my kids. I think that there are many times when kids make decisions naturally." Lance agreed, stating that Robert Stahl's cognitive approach "is a natural progression that we go through as humans." Teachers connect the natural process of building knowledge with the decision-making phases identified by Stahl. R.J. continued, establishing a direct link between the cognitive phases of decision-making and social studies content:

> So, without sounding political, we deal extensively with current events and issues [in history]. One of the most important topics for the last couple of months has been Ukraine and Russia. And I was reading through the steps [decision-making], and I thought those are important [to the teacher] to direct lessons because sometimes they [students] are apathetic, tired, or have not developed critical thinking skills yet.

Teachers highlighted current societal issues related to the lack of inquiry-based decision-making opportunities. All seven teachers identified concerns in today's society, citing the absence of a decision-making curriculum as a primary factor. In this context, Rebecca, with twenty-five years of experience in high school social studies, shared:

> I think that is important [decision-making materials] also because, you know, education is very politicized right now. You know, there is all this talk about like what content is in the textbooks, and we are going to, you know, ban these books from libraries, and people want this sort of total control for their political agendas. I want to think that, at some point, we will get a generation of human beings who can move beyond this kind of attack culture that we see in politics. And that is another reason I want people to feel empowered to express their opinions in a way they are respected by others- to move to more productive discourse.

Theme 2: The Hindrance of Standards and Accountability detracts from Teaching

When asked about integrating decision-making materials into the social studies classroom, five of seven teachers cited issues with lesson compatibility. They viewed standards-based curriculum requirements and accountability measures as the primary obstacles to planning and implementing decision-making and inquiry-based activities. Jack Schneider and Andres Saultz (2020) describe the tension between curricular freedom and state standards-based accountability as understandable. The state's education department cannot relinquish control to local schools without losing its authority to set standards and accountability measures. Meanwhile, teachers like Laura seek freedom from standards-based curricula and the pressures of high-stakes testing.

> One of the things that bugged me the most is that I do not feel that within our

> school system if we could start with the decision-making process on the teacher a level that would make things easier for the kids. So that we could get more in-depth in things, for instance, in All Quiet on the Western Front, they have to teach background information on who is on whose side, the allies, and the axis. It is kind of like social studies could cover that if it was in line with mine, and then you. I could be sitting here because of some language issues we might encounter.

Laura discussed the mismatch among state standards across subjects. One challenge she mentioned was the difficulty teachers face when working across cultures due to the state's progression standards. She said that English and social studies teachers should be able to work together by sharing literacy and historical content, which could be achieved by realigning the standards.

Another common obstacle mentioned by participants is the pressure of end-of-course testing. All seven participants said they could not realign standards across all decision-making curricula because they needed to cover specific topics by scheduled dates to ensure they were prepared for assessments. Rebecca, a World History social studies teacher, expressed concerns about the erosion of teacher pedagogy caused by standards-based accountability measures.

> In my perfect classroom world, I would not have to cover so many topics and could incorporate several of the decision-making activities provided. I would like it if there were fewer things I had to teach, and I could teach each topic in more depth with inquiry-based activities where students could face a problem or dilemma and find plausible solutions and outcomes. I want students to investigate more primary sources rather than focus on multiple-choice-oriented tests. I think many teachers are intimidated by district and End of Course tests and are reluctant to incorporate decision-making materials for fear that all content is not covered.

Research by Virtue, Buchanan, and Vogler (2012) acknowledged that standards-based accountability challenges social studies teachers to find time and suitable materials to incorporate creativity and inquiry. However, they suggested that teachers should dedicate instructional time to implementing inquiry-based activities or lessons within a unit after completing high-stakes testing.

Theme 3: Social Studies on the Back Burner to Other Subjects

Because all participants teach exclusively within social studies departments at their schools, their responses focused on social studies instruction, as well as their perceptions of and the image of social studies within their schools, districts, and the state departments of education. According to Neal Houser (1994), the term "back burner" is a metaphor for social studies, which have been undervalued and underrepresented. He writes, "Social studies are one of the things you put on the back burner – every other subject, including mathematics and science, are focused on first" (p.17). Six of the seven participants agreed with Neal Houser's position. Four teachers commented on the perceived second-rate status of social studies among curricular subjects. Lance, who works in a predominantly rural area, expressed disdain for his social studies experience, likening it to a redheaded stepchild, and justified his view by saying:

> "I mean because in every school that I have taught, you know what they spend all the money for—math materials and English ELA materials and even science materials. The administration must forget we are also state tested, and they need activities and lessons that encourage critical thinking."

Sarah, a seven-year veteran from a large suburban school district, expressed her frustration over the lack of resources and materials.

> It is frustrating because we do not have an updated United States history text for my students to use. The worksheets that align with the textbook are not good at either making students think or creating discussion like the content-centered activities provided. Even using these lessons would work great as supplemental materials, but my school has none. It seems the focus is on everything except social studies. If I wanted to purchase a curriculum or materials like content-centered activities to use with the text, it would not be easy to do. If it is not math, science, or even English, then you [the subject] is last in line for anything.

Lintner (2006) and VanFossen (2005) noted that teachers and students find it difficult to see the importance of maintaining fidelity in the classroom when relevant materials and curriculum are missing, while school budgets prioritize funding for additional resources related to numeracy and literacy.

Discussion

Overall, the teachers interviewed unanimously identified three main themes in modern secondary social studies classrooms. These concerns are interconnected and influence one another. The first theme emphasized the importance of decision-making materials in the classroom. Modern educational standards require teachers to include inquiry-based opportunities; however, barriers such as standardized testing and accountability (the second theme) often leave teachers unfamiliar with inquiry and struggling to create these learning experiences. Teachers admitted they did not understand the term inquiry or how to implement inquiry-based learning before participating in the study. However, after exposure to values-based decision-making materials that support inquiry, they found Robert Stahl's decision-making resources simple and effective for integrating inquiry into social studies lessons. Teachers consistently expressed a desire for more student-centered opportunities to facilitate inquiry-based decision-making. This goal of providing decision-making opportunities for all learners aligns with Robert Stahl's core concept of the four decision-making phases.

Additionally, teachers noted a connection between standards-based accountability and perceived authority and control in American classrooms. Tested subjects tend to be prioritized when selecting textbooks and curriculum materials. In this context, teachers are typically provided with pacing guides that serve as timelines for meeting standards; this type of accountability pressures teachers to adhere to time limits during testing periods. Our interviewees described this as a barrier to their creativity and autonomy—restrictions on their ability to pursue inquiry due to these accountability measures.

While teachers agreed that standards-based accountability hampers their ability to develop and incorporate inquiry-based decision-making, they also noted that social studies is often considered a second-tier subject within the curriculum. Regarding the uneven distribution of content, secondary social studies teachers recognize that a lack of teaching materials and resources directly affects the effectiveness of their instruction. Both teachers and students find it challenging to maintain fidelity in social studies education under these conditions.

Conclusion

During this study, the researchers examined perceptions of current instructional practices, inquiry-based learning, and the influence of state and national standards on teaching. All seven teachers interviewed considered inquiry-based learning vital to social studies instruction. Standards-based accountability was viewed as a barrier to planning inquiry-based learning opportunities. Finally, teacher participants shared their perceived limitations and restrictions when implementing inquiry-based instruction in the classroom, as well as the overall lack of emphasis on social studies education.

Overall, the researchers believe that the teacher participants in this study want to create more opportunities to include decision-making materials in their daily lessons. The teachers mentioned plans to incorporate decision-making activities, including those provided in future lessons. Additionally, five of the seven teachers interviewed requested more reading by Robert Stahl (1997) to help them develop decision-making and inquiry-based activities. While the researchers see that social studies teachers understand the importance of covering a wide range and depth of content, the teacher participants in this study unanimously felt a general disregard from state and local officials for social studies. The current research revealed that teachers perceive high school social studies as often deprioritized and ignored, despite their usefulness. It was also suggested that a lack of funding, classroom resources, and the high-stakes focus on mathematics and English Language Arts directly compete with social studies.

References

Agarwal, P., & Roediger, H. (2019, January). Lessons for Learning: How Cognitive Psychology Informs Classroom Practice. *Phi Delta Kappan*, *100*(4), 1-5. https://doi.org/10.1177/003172178815666.

Ahmed, I. (2017) The Third World Perspective on the Cold War: Making Curriculum and Pedagogy Relevant in the History classroom. *Journal of International Social Studies. 7,* 121-135.

Balakrishnan, V. (2009) Teaching moral education in secondary schools using real-life dilemmas.*I-Manager's Journal on Educational Psychology. 2*(3), 14-20.

Berman, J. (1986). *Law and Revolution: The Formation of the Western Legal Tradition.* Cambridge: Harvard Press.

Boddington, T. (1980), ' Empathy and the teaching of history,' *British Journal of EducationalStudies, 28* (1), pp. 13-19.

Casteel, J., & Stahl, R. (1997). *Doorways to decision making*. Prufrock PR.

Conklin, T., & Boulamatsi, A. (2020, October). Decision-making: The process is the content in an experience-based classroom. *Decision Sciences Journal of Innovative Education*, *18*(4), 2-18. https://doi.org/10/1111dsji.12224.

Cowgill, D., & Waring, S. (2017) Historical thinking: An evaluation of students and teacher ability to analyze sources. *Journal of Social Studies Research. 8*(1), 115-145.

Dunn, R. (1983). Learning style and its relation to exceptionality at both ends of the spectrum. Exceptional Children, 49, 496–506.

Houser, N. (1994). *Social Studies on the backburner: Views from the field*. University of Delaware. Retrieved May 13, 2022, from https://files.eric.ed.gov/fulltext/ED381461.pdf.

Lee, P.J and Ashby, R. (2001), 'Empathy, perspective taking, and rational understanding,' in O.L.Davis, Jr., S.J. Foster and E.A. Yeager (Eds), *Historical empathy and perspective taking in social studies,* Boulder: Rowman and Littlefield.

Lennon, S.; Byford, J. & Cox, J.T. (2015). An Ethical Exercise for the Social Studies Classroom: The Trolley Dilemma. *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas*, 6(88), 178-181.

Levin, M., Newmann, F.M., & Oliver, D.W. (1969). *A law and social science curriculum based on the analysis of public issues.* Washington, DC: Department of Health, Education, and Welfare.

Lintner, T. (2006). Social studies (still) on the back burner: Perceptions and practices of K-5 social studies instruction. *Journal of Social Studies Research*, *30*(1), 3-8.

Merriam, S., & Grenier, R.S. (2019) *Qualitative Research in Practice: Examples for Discussion and Analysis.* 2nd Edition. Wiley, San Francisco.

National Council for the Social Studies (NCSS). (2019). *College, career, and civics life (C3) framework for social studies state standards: Guidance for enhancing the rigor of k-12 civics, economics, geography, and history*. Silver Spring, MD: National Council for the Social Studies.

Pham, L.T., & Hayden, M. (2019) Research in Vietnam: The experience of the humanities and social sciences. *Journal of International and Comparative Education. 8*(1), 27-40.

Popham, J. (2006, September 15). *Defining and enhancing formative assessment*. Assessment for Learning. Retrieved November 3, 2021, from http://www.brjonesphd.com/uploads/1/6/9/4/16946150/defining_formative_assessment_popham.pdf.

Portal, C. (1987), *The history of curriculum for teachers,* London: Palmer Press.

Pratt, A., & Richard, A. (2019, November 26). Improving education in rural Tennessee schools in *SCORE: State Collaborative on Reforming Education*. Retrieved February 22, 2022, from https://tnscore.org/improving-education-in-rural-tennessee-schools/.

Rappoport, A. (2006) Least known to Americans: Content materials about the Soviet Union in the 1940s and 1950s. *The Social Studies. 97*(2), 56-61.

Rugg, H. (1927). *The foundations of curriculum-making: Twenty-sixth yearbook of the National Society for the Study of Education*. Bloomington, IL: Public School Publishing.

Schlene, V.J. (1996) *Teaching About Vietnam and the Vietnam War.* ERIC Clearinghouse for Social Studies/Social Science Education (ED411175).

Schneider, J., & Saultz, A. (2020). Authority and control: The tension at the heart of standards-based accountability. *Harvard Educational Review*, *90*(3), 419-445. https://doi.org/10.17763/1943-5045-90.3.419.

Sheurman, G. (2018). *Constructivism and the new social studies*. Information Age Publishing.

Stahl, R. (1976, October). Values morals education: A synthesis model. *ERIC Clearinghouse of Teacher Education*. Retrieved May 3, 2021, from https://files.eric.ed.gov/fulltext/ED129718.pdf.

Stahl, R. (1978). Moral dilemmas value sheets: Writing for content centered social studies classrooms. *Southeastern Region of the National Council for the Social Studies*, 2-20.

Stahl, R. (1979). *The effects of values dilemma activities on the content retention and attitudes of social studies students: An empirical investigation based upon the Casteel-Stahl approach to values education*. edited by Arizona State University, Portland, National Council for the Social Studies. Retrieved May, 4, 2021, from https://files.eric.ed.gov/fulltext/ED196748.pdf.

Stahl, R. (1995). *Society and science: Decision-making episodes for exploring society, science and technology*. Addison Wesley Publishing.

Stake, R. E. (2017). *The art of case study research*. SAGE Publications.

Strauss, A., Schatzman, L., Butcher, R., & Sabshin, M. (1981). *Psychiatric ideologies and institutions* (2nd ed.). New York: Wiley.

VanFossen, P. J. (2005). Reading and math take so much of the time: An overview of social studies instruction in elementary classrooms in Indiana. *Theory and Research in Social Education*, *33*(3), 376-403.

Virtue, D., Buchanan, A., & Vogler, K. (2012). Digging postholes adds depth and authenticity to a shallow curriculum. *Social Studies*, *103*(6), 247-251. https://doi.org/10.1080/00377996.2011.630699.

White, R. (1994) 'An alternative approach to teaching history.' *OAH Magazine of History, 8*(2), 58-60.

Yin, R.K. (2017). *Case study research and applications: Design and methods* (6th ed). 50-83. Los Angeles, CA: Sage.

Group Presentations After College: Are Students Prepared?

Andres Zapata
University of Tampa

Dr. Chris Gurrie
University of Tampa

Abstract
Group projects are a way of life for college students. From team presentations to research cohorts, students often find themselves working together. Yet, how much like the "real world" is this practice? Specifically, in multi-person presentations in the live format, does college really prepare students for careers where presentations take place? This mixed-method study surveyed college graduates about their experience in team public speaking professional spaces and asked their opinions about how their college training prepared them for work. Thematic analysis was used to compare the results to suggested literature review best practices. Many respondents felt college somewhat prepared them for team presentations at work, but often it was more about interpersonal relationships than the actual presentations themselves.

Keywords: Communication Education, Organizational Communication, Teams

Introduction and Literature Review

Before understanding whether group presentations truly prepare students for professional environments, it is important to first understand the problem they are supposed to solve. Across different industries and countries, researchers continue to find a major gap between the skills employers expect from recent graduates and the skills graduates bring into the workforce. While many students leave college feeling confident in their communication and teamwork abilities, employers often disagree with those self-assessments. A nationally representative survey of U.S. business executives found that although nearly 80% of recent graduates believed they were highly skilled in teamwork and communication, fewer than 40% of employers shared that opinion (Hoque et al., 2023). This disconnect is important because it suggests that many graduates are not only underprepared for professional collaboration, but also unaware of how unprepared they may be (de Prada Creo et al., 2022).

Similar concerns have also been found outside the United States. Research conducted in Australia identified self-management, oral communication, teamwork, cognitive skills, and creativity as some of the most consistently lacking competencies among recent graduates entering the workforce (Prikshat et al., 2019). The fact that these concerns appear across multiple countries and professional environments suggests that this is not simply an isolated issue within one educational system. Instead, it points toward a larger structural disconnect between what higher education teaches and what professional environments expect from students once they graduate.

Even though researchers generally agree that a skills gap exists between college graduates and workplace expectations, there is far less agreement on what that gap actually looks like. Suleman (2016) found that employers and educators consistently emphasize the importance of relational skills such as interpersonal communication, teamwork, and collaboration, but there is little consistency in how those skills are defined or measured across studies. In some cases, employers report that graduates struggle with teamwork and communication while still performing adequately in technical or digital skills. In others, oral presentation abilities are considered essential, even when students continue to perform strongly on traditional written assignments (Suleman, 2016).

This lack of consistency becomes especially important when discussing group presentations in higher education. If educators and employers do not fully agree on what effective “working-world communication” looks like, it becomes difficult to determine whether group presentations are truly helping students develop those skills. In many ways, the challenge is not simply whether students are communicating, but whether they are practicing the kinds of communication that professional environments demand.

Some of the strongest evidence supporting group presentations comes from research focused on work-integrated and project-based learning. Studies suggest that students benefit the most when presentations are connected to authentic professional experiences rather than traditional classroom assignments. Twyford and Dean (2023), for example, examined an industry-led accounting simulation where students presented financial recommendations to professionals acting as clients. Students reported significant improvements in teamwork, interpersonal communication, professional confidence, and project management skills after completing the assignment (Twyford & Dean, 2023). Many participants also emphasized how transferable the experience felt to future workplace situations, especially because the assignment involved real feedback, real audiences, and realistic professional expectations (Twyford & Dean, 2023). One student even described the experience as “much needed professional experience in how I will conduct myself in the workplace” (Twyford & Dean, 2023, p. 19).

Similar findings appear in project-based learning research. Naseer et al. (2025) found that students participating in industry-integrated collaborative projects demonstrated significantly stronger improvements in employability skills than students in more traditional learning environments. The largest gains were connected to teamwork, communication, and problem-solving, while presentation-based project pitches received some of the highest student experience ratings. Overall, the study found that more immersive and structured collaborative experiences tended to produce stronger professional skill development.

At the same time, research also suggests that group presentations are far less effective when they lack authentic context, structured guidance, or meaningful reflection. A study of 615 university students in Spain found major differences between the skills students believed they had developed through group work and the skills employers actually valued (De Prada Creo et al., 2022). While students often identified communication and teamwork as strengths, they were less likely to recognize weaknesses in areas such as research synthesis, problem framing, and professional presentation of information (De Prada Creo et al., 2022). Ornellas et al. (2019) argue that authentic learning environments, especially those involving real-world relevance, collaboration, reflection, and professional feedback, are essential for developing transferable employability skills. Without those elements, group presentations may encourage collaboration on the surface while failing to develop the deeper communication and analytical skills expected in professional settings.

Although the literature still provides limited evidence about what skills directly transfer from classroom presentations into workplace environments, industry engagement consistently appears to strengthen the learning experience. Naseer et al. (2025) found that industry mentorship was rated as the most valuable form of engagement for developing workplace-relevant communication skills, while Twyford and Dean (2023) similarly found that students viewed direct interaction with professionals as one of the most meaningful aspects of the presentation experience.

Research suggests that many communication and speech programs struggle to design group work in ways that truly support collaborative learning. This is not necessarily because instructors do not care about the process, but because effective group work requires a level of structure and guidance that is often missing in traditional classroom settings. Fittipaldi (2020) describes the most common approach as a “sink or swim” model, where instructors assign the project, set a deadline, and then deal with issues like unequal participation, social loafing, and group conflict only after problems arise. After reviewing more than 46 empirical studies on group project management, Fittipaldi (2020) identified several evidence-based practices that improve group work outcomes, including smaller group sizes, team training, team contracts, continued instructor involvement, and both formative and summative peer evaluations. However, these strategies are rarely implemented consistently in undergraduate communication courses.

One of the biggest weaknesses in current group presentation design is the assumption that students naturally know how to collaborate. Research shows that collaboration itself often needs to be taught. Prichard et al. (2006) found that students who received team-skills training before beginning collaborative work performed significantly better than students who did not receive that training. However, those improvements disappeared once the groups were reorganized between semesters, suggesting that long-term collaboration and group stability also play an important role in skill development (Prichard et al., 2006). Similarly, Sjolie et al. (2021) found that both team-skills instruction and active instructor facilitation improved collaborative learning outcomes by helping students develop a shared understanding of how teamwork functions.

Group size also appears to have a major impact on the success of collaborative assignments. Research consistently shows that smaller groups of four or five students tend to communicate more effectively, experience fewer conflicts, and distribute work more fairly than larger teams (Fittipaldi, 2020). In contrast, teams with more than six members often divide into smaller subgroups that exclude certain students from meaningful participation (Kimpton & Maynard, 2024). Despite these findings, many courses still organize groups based on convenience rather than evidence-based practices.

Another major issue involves accountability. Loh and Ang (2020) argue that successful cooperative learning depends on structures that clearly separate individual contributions from overall group outcomes. Peer assessment is one of the most common ways instructors attempt to address this issue. Fittipaldi (2020) found that stronger peer evaluation systems can reduce free-riding and encourage more active participation, especially when evaluations are completed throughout the project rather than only at the end. However, peer assessment is only effective when students are properly trained on how to evaluate one another fairly. Without that guidance, peer reviews can become unreliable due to personal bias or inconsistent standards (Fittipaldi, 2020).

Researchers also argue that the problem extends beyond individual assignments and reflects broader issues within higher education curriculum design. Boud and Bearman (2022) explain that traditional grading systems often work against the goals of collaboration because students are simultaneously expected to work together while competing for individual grades. Karimi and Piña (2021) similarly argue that communication and teamwork skills should not be treated as one-time classroom activities, but as long-term learning outcomes integrated throughout an entire degree program. In other words, a single group presentation in one speech class is unlikely to fully prepare students for professional collaboration after graduation.

Cultural and demographic dynamics also shape how students experience group work. Loh and Ang (2020) note that students from different cultural backgrounds may approach communication, authority, and group harmony differently, which can directly affect participation and peer feedback. Kimpton and Maynard (2024) additionally found that women in male-dominated engineering teams were frequently pushed into non-technical or supportive roles rather than leadership positions, even when they demonstrated strong technical abilities. These kinds of inequalities not only create unfair group experiences but also limit students' opportunities to develop the professional communication skills group presentations are intended to teach.

Even with the structural problems surrounding group work in higher education, research still shows that group presentations help students develop several important professional communication skills. One of the most consistent benefits is interpersonal communication within collaborative environments. Loh and Ang (2020) found that cooperative learning improves students' ability to manage disagreement, interpret social dynamics, and work through conflict constructively. Twyford and Dean (2023) similarly found that students developed a stronger understanding of teamwork by learning how to support one another's strengths and collaborate more effectively in group settings.

Group presentations also appear to strengthen skills related to time management and project coordination. Arsenis et al. (2021) found that students frequently identified meeting coordination, task delegation, and role management as major outcomes of collaborative assignments. Kimpton and Maynard (2024) similarly noted that students consistently viewed time management as a key factor in successful teamwork.

Research also suggests that presentation delivery skills improve most when structured feedback is included. Park and Park (2011) found that workshops involving peer feedback, self-analysis, and model presentations significantly improved students' oral presentation abilities. Likewise, Murillo-Zamorano and Montanero (2018) found that rubric-based peer assessment improved oral presentation performance more effectively than instructor feedback alone. Their findings also showed strong alignment between peer and instructor evaluations, suggesting that structured rubrics can produce reliable and meaningful feedback (Murillo-Zamorano & Montanero, 2018).

Personal accountability is another commonly reported outcome of group presentations. Twyford and Dean (2023) found that students became more aware of their responsibility to teammates throughout collaborative projects, while Fittipaldi (2020) found that team charters improved participation, task cohesion, and overall satisfaction with group work experiences.

However, the literature also suggests that group presentations are less successful at developing higher-level professional communication skills such as persuasive information synthesis, audience adaptation, and executive presence. Arsenis et al. (2021) found that although students recognized improvements in communication and teamwork, they were less likely to identify growth in research synthesis or professional presentation abilities. This may be because many classroom presentations still fail to fully replicate the conditions and expectations of real workplace communication.

Although the existing research on group presentations is extensive, several important gaps still limit how clearly scholars can determine whether these assignments truly prepare students for professional environments.

One major limitation is the lack of long-term research following students after graduation. Most studies measure outcomes immediately after the assignment or at the end of a course, rather than examining whether communication and collaboration skills are actually retained and applied in workplace settings over time. Tight (2023) notes that employability research more broadly often depends on student self-reports and employer perceptions instead of long-term behavioral evidence.

Another major gap is the lack of research focused specifically on communication and speech courses. Most studies on group presentations come from fields such as business, engineering, or health sciences, while communication-focused programs remain surprisingly understudied. Because of this, it is difficult to determine whether speech and communication curricula produce stronger collaborative presentation outcomes than other disciplines.

The growing shift toward virtual presentations is also underexplored in current literature. Lu and Smiles (2022) found that online collaborative learning develops a unique set of communication skills, while

Kimpton and Maynard (2024) noted that communication and time-management challenges become even more complicated in virtual team environments. However, very little research examines how online academic presentations prepare students for remote or hybrid workplace communication, despite the increasing normalization of virtual professional environments.

Research on instructor preparation is similarly limited. Although studies frequently describe the "sink or swim" approach to group work (Fittipaldi, 2020), there is still little understanding of why many instructors continue to rely on it. Loh and Ang (2020) suggest that barriers such as limited time, large class sizes, and lack of institutional support often make meaningful cooperative learning difficult to implement consistently.

Finally, much of the literature tends to treat all forms of "group work" as interchangeable. Group research projects, video assignments, collaborative presentations, and professional simulations are often discussed together even though they involve different communication demands and learning outcomes. As a result, it remains difficult to isolate what group presentations specifically contribute to professional communication development compared to collaborative assignments more broadly.

Purpose and Research Questions

After reviewing the literature and understanding what gaps may be in the literature the author team wanted to know:

RQ: Are group presentations preparing students for working group presentation scenarios?

RQ: Are communication programs and/or speech classes assigning group work correctly?

RQ: What presentation skills in group presentations to college get right?

To better understand this the team conducted a thematic analysis to identify how this work could better prepare curriculum specialists, communication scholars, and other stakeholders preparing to administer and evaluate group presentations in college and after. What follows are the sections of methods, results, discussions, limitations, and future planning for this work.

Method

This primary research was conducted using a snowball sampling of post-college professionals for a convenience sample through online networks and social media outlets. The survey instrument (Appendix A). The questionnaire asked the following questions:

1) Please describe as a working professional, how are group PRESENTATIONS different for you now, than they were in college

2) Now that you're out of college, what are the biggest differences between group presentations in university and now?

3) What advice would you give a college student about working in group presentations as a graduated professional employee/worker?

4) How did group presentations in college prepare you for team presentations after college?

5) Anything else you want to tell the research team about group projects or presentations as a working professional versus when you were in college?

This was an IRB approved study application number 52113349 confirmed November 2025. The researchers reside in and work at a university in the Southeastern part of the United States. The method used to interpret the data is Creswell's Thematic Analysis for qualitative data (2008). Described for understanding, the use of themes is a way to analyze qualitative data. Themes are similar codes aggregated together to form a major idea in the dataset. Like codes, themes have labels that typically consist of no more than two to four words (Creswell, 2008).

The open-ended responses were coded for key phrases and terms. They were then placed into the typical categories of: ordinary themes, unexpected themes, hard-to-classify themes, and major and minor themes. Specifically, the responses are printed, analyzed for thematic patterns, and then grouped together (themes). Layering comes to provide rigor when explaining subsets within the themes (Creswell, 2008). The results will demonstrate these so readers can see the varied data themes and how they relate to answering the questions in the abstract and in the applied with the purpose of adding to the body of literature on this topic of group presentations inside and outside the college classroom.

Results

Across the four open-ended questions there were 77 responses that yielded data that were coded into themes. When analyzed, key terms, words, and phrases that were repeated by respondents were noted by the researchers. Then, based on frequency, the words and phrases were conflated to create the themes. The most common response was related to *weight, balance of work, or distribution of labor* related to presenting a group presentation. This was labeled the ordinary theme. Within that category of ordinary was the major theme of *effort*. The researchers delineated effort differently from division of labor because this implied that some people may try harder than others–for whatever reason. The minor theme was *design or text* of the visuals related to the PowerPoint. While unexpected, this was labeled as such because that marker was reserved for *audienc*e. The unexpected theme for this dataset was the group's *audience*. Several respondents stated they changed their presentation behavior based on to whom they were presenting; suggesting that if it was just the college professor and for a grade they did not care very much. However, in the professional world where money and work were on the line, they changed their level of presentation caliber.

Table 1

Theme	Words/Phrases	Frequency/=n
Ordinary Theme	Weight, balance, distribution of labor	26
Major Theme	Effort	16
Minor Theme	Design/text/PowerPoint	6
Unexpected Theme	Audience	10

These four themes were evident across all questions and regardless if the question inquired about college or the professional world. Respondents noted that distribution of labor, effort, slide design, and audience were important no matter the context of the group presentation. For readers unfamiliar with Creswell's theming, adding multiple perspectives to the reported data means a researcher provides varying viewpoints which help convey the complexity of the phenomenon (Creswell, 2008). To do this, below are sample phrases from respondents' answers to the relevant questions that were coded into themes.

Q1. Please describe in as much detail…what were group presentations like for you in college?
Ordinary

Participants did not carry their own weight

Some people were missing in action

One person took the lead

Major

Easy to collaborate with

Hard to reach

It was definitely easier with the group when we practiced

Minor

None

Unexpected

The professor was often the only active participant

Presentations were an act of pure information exchange letting the professor know…

Fun or frustrating depending on the people

Q2. What are the similarities between your current professional group/team experience and your college experience?

Ordinary

People tend to not change over time based on their respective privilege

Some would rather stay busy with their planning and brainstorming than actually present

Almost nothing [is different from college]

Major

Everything! I was prepared for the real world by college.

Some peers prepared a lot

Some people may disappoint you

Minor

Being clear and concise is important [in Email and discussions]

Unexpected

None

Q3. What are the differences between college group presentations and your current professional group/team presentations?

Ordinary

Slackers still exist

There is far more accountability at work

Everyone has skin in the game

Major

Less tolerance for people not being prepared

Minor

Very rare planned PowerPoint to large scale clients at their facility

Today it's more status updates and pitches and academia more toward racaps of projects

More customer stories and more PowerPoint but more pictures and less words

Unexpected

More reality based not speculative [like college]

More overlap with projects and direct coworkers

Q4. Is there anything else you want to tell the researchers about group presentations in the workplace?

Ordinary

Carry your load, communicate often, be present and do your best to be a professional at college and work

Major

I've done interview prep with hundreds of people throughout my career. There are always folks who are extremely uncomfortable with public speaking, and folks who refuse to prepare because they claim to be better at "winging it" versus preparing. I do think some people are trainable and can improve. But there are others who simply cannot perform well in a group setting.

Being able to communicate ideas clearly and keep things concise

Learning to read the room and adjust your tone and focus will make a huge difference later on.

Minor

None

Unexpected

Practice social skills and learning how to work/communicate in small groups

Being a good presenter is massive for your career trajectory. Knowing what to present to what audience is critical. Your audience usually has no idea what you are talking about so you need to build foundations quickly while presenting.

One thing I would tell students is that strong presentation skills and teamwork matter way more than they might think. Being able to communicate ideas clearly, keep things concise, and handle questions confidently is a big deal in the real word. It's also important to learn how to adapt your message to your audience. In college you present to classmates and professors, but in the workplace your audience could be executives, clients, or technical experts. Learning to read the room and adjust your tone and focus will make a huge difference later on.

Discussion

Circling back to the research questions, the researchers wanted to better understand:

RQ: Are group presentations preparing students for working group presentation scenarios?
RQ: Are communication programs and/or speech classes assigning group work correctly?
RQ: What presentation skills in group presentations to college get right?

It seems the questions from the instrument better compare college versus the professional world rather than directly answer the questions based on the themes. It seems in some cases that the context is what affects who carries weight, who puts in effort, or who the audience may be. So, inherently one could answer "no" that college presentations do not prepare students for their working presentation experience.

On the other hand, across all questions there were common themes like division of labor, effort, audience, and design. This would indicate that student/presenter behavior stays the same despite the context in which the presentation is happening. Therefore, one could argue that group presentation assignments are currently just fine unless a college professor wanted to target the roles, responsibilities, and effort of group members both during the pre-work and the actual presentation. This, however, may be more discursive when studied in a micro-context of discipline in college compared to discipline of professional working respondent. For example, are public relations professors assigning the right group projects for those who will become public relations professors? The current study simply asked if respondents felt "college" (at-large) was similar, or prepared them for the professional world.

Finally, the one thing it seems colleges and curricula are getting "right" is the assigning of group work. While some respondents stated they make few presentations annually, not one response across the 77 stated something close to "I don't do group presentations," or, "there are no group presentations in my working world." This may create a question for further inquiry about the frequency of assigned group presentations in college versus those of the professional world. Perhaps sometimes when students anecdotally complain about group presentations, they may be tired of the frequency over four courses, several subject matters, different or unknown groupmates, and more. Whereas, the data from the current study indicate professionals know their co-presenters and group members and often they are experts in the subject matter.

Similar to paying closer attention to the roles presenters play in a group, this too could call for attention to be paid to the number and frequency of college group projects and presentations. This, then, may put this work more closely aligned with studies in curriculum and instruction or teaching, rather than speech communication or business. In summary, it seems that college curriculum is getting it “right” with the group presentations but the way the work is divided remains a problem from college to boardroom, and possibly a problem colleges cannot change.

Limitations

In synthesizing this work, sometimes it is blurry deciphering what may be limitations and what might be future opportunities. Perhaps the snowball convenience sample procedure was a limitation because it recruited participants in the researchers’ networks which may have seen and unseen similarities to each other and the researchers. On the flipside of that, this work can be seen as a preliminary study to larger research that spends more time sampling different people from different fields. In fact, the instrument could be used for various fields or majors and even cross tabulated between those specific fields in the college and professional worlds.

While markers were collected to make sure everyone was beyond college and had participated in some college courses, the work did not collect demographic information in a way that could divide if women felt one way and men another. In the future taking advantage of the Qualrics forms so this could be spliced easier would be better for demographic analysis if relevant.

Future

Conducting this study only opened up more questions for the researchers. For example, should speech and communication faculty members focus on [teaching] the roles and channels of group presentations, whereas discipline specific group content should focus on audience and purpose? Can one exist without the other? Is it really a package presentation of content and delivery? Future studies could add to the social science literature of group communication and presentation behavior, as well as curriculum and instruction work. Moreover, this study can be the foundation of teaching and learning talks at the researchers’ institutions and beyond. If any reader wants to reach out for ideas or partner in the future, that collaboration could better use this information with the purpose of helping others.

Conclusion

Group projects and presentations are a modern way of life for college students. This study was created to better understand if colleges and universities are “getting it right” with group presentations in college when preparing students for the professional world. While there are countless ways to assign and deliver group presentations, the researchers asked participants to think generally about their college experiences and then relate them to their current working practice. Decidedly, the respondents had no problems providing their experiences and many of them were similar. Using Creswell's thematic analysis to understand the data and then interpreting those results for the discussion above; this current study concludes that colleges are getting it "right" with assigning the presentations themselves, but more work can be done in understanding roles, responsibilities, audience, and content.

References

Arsenis, P., Flores, M., & Flores, O. (2021). Enhancing graduate employability skills and student engagement through group video assessment. International Journal of Management Education, 19(3), 100563. https://doi.org/10.1080/02602938.2021.1897086

Boud, D., & Bearman, M. (2022). The assessment challenge of social and collaborative learning in higher education. Studies in Philosophy and Education, 41(5), 539–553. https://doi.org/10.1080/00131857.2022.2114346

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Pearson Merrill Prentice Hall.

De Prada Creo, E., Mareque, M., & Portela-Pino, I. (2022). Teamwork skills in higher education: Is university training contributing to their mastery? Psicologia: Reflexão e Crítica, 35(5). https://doi.org/10.1186/s41155-022-00207-1

Donald, W. E., & Ford, R. (2022). Fostering social mobility and employability: The case for peer learning. Teaching in Higher Education, 29(3), 671–685. https://doi.org/10.1080/13562517.2022.2145467

Fittipaldi, C. J. (2020). Managing the dynamics of group projects in higher education: Best practices suggested by empirical research. Universal Journal of Educational Research, 8(6), 2352–2363. http://www.hrpub.org/journals/article_info.php?aid=9141

Hoque, K. E., Abdullah, A. G. K., Joarder, M. H. R., & Islam, R. (2023). The desired employability skills and work readiness of graduates. SAGE Open. https://doi.org/10.1177/09504222221149850

Karimi, S., & Piña, A. A. (2021). Strategically addressing the soft skills gap among STEM undergraduates. Journal of STEM Education: Innovations and Research, 22(1). https://j-stem.net/index.php/jstem/article/view/99

Kimpton, A. N., & Maynard, N. (2024). Factors shaping teamwork skills development in tertiary engineering education: A systematic literature review. European Journal of Engineering Education.https://doi.org/10.1080/03043797.2024.2357343

Loh, C. Y. R., & Ang, C. F. (2020). Unravelling cooperative learning in higher education. Research in Social Sciences and Technology, 5(2), 100–124. https://doi.org/10.46303/ressat.05.02.5

Lu, J., & Smiles, R. (2022). The role of collaborative learning in online education. International Journal of Engineering Business Management, 14. https://ijebmr.com/uploads/pdf/archivepdf/2022/IJEBMR_968.pdf

Murillo-Zamorano, L. R., & Montanero, M. (2018). Oral presentations in higher education: A comparison of the impact of peer and teacher feedback. Assessment & Evaluation in Higher Education, 43(1), 138–150. https://doi.org/10.1080/02602938.2017.1303032

Naseer, S., Bhatti, A. M., Majeed, F., & Khan, M. N. (2025). Project-based learning framework integrating industry collaboration to enhance student future readiness in higher education. Scientific Reports, 15.https://doi.org/10.1038/s41598-025-10385-4

Ornellas, A., Fàbregues, S., Fernández-Fernández, S., & Ferreria-Vorkapic, C. (2019). Enhancing graduates' employability skills through authentic learning approaches. Higher Education, Skills and Work-Based Learning, 9(1), 107–120. https://doi.org/10.1108/HESWBL-04-2018-0049

Park, K. H., & Park, I. B. (2011). The development and effects of a presentation skill improvement program for medical school students. Korean Journal of Medical Education, 23(4), 285–294. https://doi.org/10.3946/kjme.2011.23.4.285

Prichard, J., Bizo, L. A., & Stratford, R. (2006). The educational impact of team-skills training: Preparing students to work in groups. British Journal of Educational Psychology, 76(1), 119–140. https://doi.org/10.1348/000709904X24564

Prikshat, V., Kumar, S., & Nankervis, A. (2019). Australian graduates' work readiness – deficiencies, causes and potential solutions. Higher Education, Skills and Work-Based Learning, 10(1), 62–78. https://doi.org/10.1108/HESWBL-02-2019-0025

Ragusa, A. T., Caggiano, V., Tsecouras, A., & Brucculeri, M. C. (2022). High education and university teaching and learning processes: Soft skills. International Journal of Environmental Research and Public Health, 19(17), 10699. https://doi.org/10.3390/ijerph191710699

Salmons, J., & Wilson, L. A. (2023). Learning to collaborate, collaborating to learn. Routledge. https://doi.org/10.4324/9781003445708

Sjolie, E., Strøm-Andersen, N., Bratholm, T., & Mausethagen, S. (2021). Team-skills training and real-time facilitation as a means for developing student teachers' learning of collaboration. Teaching and Teacher Education, 105, 103399. https://doi.org/10.1016/j.tate.2021.103399

St. Louis, K. O., Myers, H. F., & Schmuldt, T. (2021). Infusing 21st century skill development into the undergraduate curriculum. Journal of Microbiology & Biology Education, 22(2). https://doi.org/10.1128/jmbe.00180-21

Suleman, F. (2016). Employability skills of higher education graduates: Little consensus on a much-discussed subject. Procedia – Social and Behavioral Sciences, 228, 169–174. https://doi.org/10.1016/j.sbspro.2016.07.025

Tight, M. (2023). Employability: A core role of higher education? European Journal of Higher Education, 14(1), 1–14. https://doi.org/10.1080/13596748.2023.2253649

Tumpa, R. J., Siddiqui, M. T. H., Hossain, M. B., & Islam, M. A. (2022). A holistic overview of studies to improve group-based assessments in higher education: A systematic literature review. Sustainability, 14(15), 9638. https://doi.org/10.3390/su14159638

Twyford, E., & Dean, B. A. (2023). Inviting students to talk the talk: Developing employability skills in accounting education through industry-led experiences. Accounting Education, 33(1), 105–131. https://doi.org/10.1080/09639284.2023.2191288

Appendix A
Survey Instrument

Professional Group/Team Presentations: Different than College?
To participate in this study participants should be working after they left (18yrs.+)/completed college; or be over 18 and professionally working while in/returning to college. The researcher is not interested in dual enrolled high school students.

1) Please describe in as much detail as you have time for, what were group PRESENTATIONS like for you in college? We are describing these as a) 2 people or more and b) a performative speech/delivery in front of an audience of any kind.

2) What are the *similarities* between your CURRENT professional group/team presentations and the college experience?

3) What are the *differences* between college group presentations and your current professional group/team presentations?

4) Is there anything else you would like to tell the researchers about group presentations in the working/professional world that would be helpful for students in college to know?

Global Engagement as a Social Science Problem
Applying the Johnson Ranking of Global Engagement (JRGE): A Case Study of a Public R1 University in the Southwestern United States

Dr. Lincoln Duane Johnson
University of Nevada, Las Vegas

Abstract

Global engagement in higher education is frequently discussed as a strategic or reputational objective, yet it is rarely examined as a measurable social science construct grounded in institutional data and theory. This paper presents a case study of a Public R1 University in the Southwestern United States using the Johnson Ranking of Global Engagement (JRGE), a globalization scorecard developed to assess how universities operationalize global engagement across teaching, research, service, and institutional infrastructure.

Drawing on social science perspectives from globalization theory, organizational sociology, and education policy, the study employs a mixed-methods design incorporating document analysis, institutional data review, and indicator-based scoring. Results illustrate how global engagement manifests unevenly across institutional units and how existing global ranking systems obscure these internal dynamics. The JRGE provides an alternative framework that emphasizes transparency, contextualization, and institutional learning rather than competitive prestige.

The paper concludes by discussing the broader implications for U.S. higher education, including how universities can use globalization metrics to guide strategic planning, improve accountability, and respond to global social, economic, and educational challenges. The JRGE is positioned as a scalable model with relevance for universities nationally and internationally.

Keywords: globalization, internationalization, ranking, higher education, evaluation, case study

Introduction

Global engagement has become a central priority for higher education institutions across the globe, reflected in expanding international partnerships, mobility programs, and globally oriented research initiatives. Despite this growth, institutions, although focused on and working towards centralized governance models, often lack the capacity to systematically measure, interpret, and manage their global activity. This has led universities worldwide to pursue more conscientious approaches to internationalization, shifting focus from quantity to quality, ethical engagement, and pedagogically grounded programs (Ledger & Kawalilak, 2018). In practice, this requires institutions to gauge, assess, and evaluate internationalization efforts using established indicators in order to determine their true global impact and influence. However, existing approaches—most notably global university rankings—privilege external reputation and aggregate outputs, offering limited insight into how global engagement is actually organized within institutions. This limitation is significant, as effective internationalization requires a "whole-of-university" approach that transcends often siloed domains such as research, teaching and learning, and engagement (Ledger & Kawalilak, 2018). As a result, universities invest heavily in global initiatives while remaining unable to clearly assess where engagement is occurring, where it is underdeveloped, and how it aligns with institutional strategy.

This limitation is not simply technical; it is organizational and structural. Global engagement in contemporary universities is typically distributed across multiple units, including academic departments, international offices, research centers, and administrative divisions. While fragmentation is common across higher education institutions, universities may adopt varying degrees of centralized coordination to address these challenges. In the case of this particular Public R1 University, internationalization functions are increasingly coordinated through a centralized global structure established within the past two years. However, the distributed nature of internationalization across colleges, research offices, and administrative units continues to create challenges for comprehensive institutional assessment. Scholars emphasize the importance of collaboration among practitioners, policymakers, and researchers, these activities are rarely integrated into a coherent measurement system (Rumbley, 2019). This produces a condition in which global engagement is simultaneously expansive and difficult to observe. As such, what appears at the institutional level as uneven or incomplete engagement often reflects not a lack of activity, but a lack of visibility. In this sense, global engagement constitutes a measurable institutional function that remains largely obscured by fragmented data systems and decentralized governance structures.

This article argues that the primary challenge facing institutions is not the absence of global engagement, but the absence of diagnostic measurement frameworks capable of rendering that engagement systematically visible, interpretable, and actionable. While the global landscape of higher education internationalization continues to shift toward more inclusive and quality-oriented models (Altbach & de Wit, 2018), institutions still lack the tools necessary to operationalize and evaluate these priorities. To address this gap, this study introduces the Johnson Ranking of Global Engagement (JRGE), a multidimensional, data-driven framework designed not as a reputational ranking, but as an internal diagnostic tool. Unlike traditional rankings, JRGE is built entirely from institutional data and is structured to identify and verify where global engagement exists, where it breaks down, and where strategic intervention is needed. In doing so, it shifts the focus from external comparison to internal analysis.

The utility of this approach is demonstrated through an application of JRGE at a large public R1 university in the southwestern United States, a large, urban, public research institution with significant but unevenly distributed global activity. Contemporary higher education institutions engage globally through a wide range of mechanisms, including student and staff mobility, internationalized curricula, cross-border programs, partnerships, and collaborative learning models (Rajkhowa, 2024). This case illustrates a central paradox: while global engagement is highly visible in certain domains—particularly in student life and partnership activity—it remains difficult to measure in others, especially in research and governance. This disconnect reveals a broader "lived–measured disjuncture," in which institutional practices exceed the capacity of existing data systems to capture them. Internationalization is most effective when it is embedded across the institution and aligned through internal resources, strategy, and infrastructure. However, without coherent measurement systems, these efforts remain fragmented and difficult to govern. As such, internationalization must be understood as a measurable institutional function that requires diagnostic tools to make it visible and actionable.

As higher education institutions expand and diversify, they have become increasingly complex organizations facing challenges across governance, administration, academic delivery, and infrastructure (Barbosa & Neves, 2020). In this context, measurement must be understood not merely as a technical exercise, but as a form of institutional governance. This study advances three contributions: First, it reconceptualizes global engagement as a distributed institutional function rather than a singular outcome. Second, it demonstrates how gaps in data are analytically meaningful, serving as indicators of organizational fragmentation and unmet strategic capacity. Third, it positions diagnostic measurement frameworks such as JRGE as essential tools for institutional decision-making, enabling universities to move from anecdotal and externally driven assessments toward systematic, data-informed governance of global engagement. In an increasingly globalized knowledge economy, internationalization is no longer a peripheral activity, but a core institutional responsibility requiring coordinated, measurable, and strategic implementation.

Conceptual Framework

The conceptual foundation of this study rests on three interrelated propositions: first, that globalization in higher education functions as an institutional property rather than a discrete activity; second, that contemporary universities exhibit structural fragmentation in how global engagement is organized; and third, that measurement is not neutral, but constitutive of what becomes visible, knowable, and governable within institutions. Internationalization in higher education has introduced new pedagogical demands, requiring teaching and learning to adapt to increasingly global and multicultural contexts (Mugano & Dorasamy, 2022). While it offers benefits such as global citizenship, cross-cultural knowledge production, and institutional growth, it also presents significant challenges, including the commodification of education, tensions in multicultural exchange, and the dominance of Western, English-language paradigms. This illustrates that internationalization is not a neutral or uniformly beneficial process, but a complex and contested institutional transformation. Global engagement, therefore, is not merely activity; it is structured, uneven, and shaped by underlying systems, making it difficult to interpret without intentional measurement. This reinforces the need for a framework such as the Johnson Ranking of Global Engagement (JRGE), as what appears as “internationalization” is often multi-layered and partially obscured without diagnostic visibility.

Global engagement is most productively understood not as an outcome or a bounded programmatic area, but as a distributed institutional function embedded across multiple domains of university activity, including academic programming, research collaboration, student mobility, partnership development, and administrative coordination. However, because these functions are often managed independently across organizational units, global engagement rarely appears as a coherent system at the institutional level. Instead, it emerges as a collection of disconnected practices that are globally oriented but not systematically integrated. In this sense, globalization is not absent within institutions; rather, it is organizationally dispersed. While international comparisons are necessary for informing policy, they require clearly defined objectives and measurable outcomes (van Gaalen et al., 2021). At the same time, internationalization reshapes core academic functions and produces simultaneous benefits and tensions, rather than linear progress. Institutional practices are therefore influenced not only by educational values, but also by market pressures and global hierarchies.

This dispersion is reinforced by the structural complexity of modern higher education institutions. Universities operate through semi-autonomous units that develop their own priorities, metrics, and reporting systems. While this structure enables specialization and local responsiveness, it also produces fragmentation in institutional knowledge about global activity. As a result, global engagement is frequently enacted without being fully aggregated or interpreted at the institutional level. The consequence is not a lack of internationalization, but a lack of organizational coherence in how it is understood and managed. This raises critical questions about the underlying values and long-term purposes of internationalization, particularly as institutions seek to balance cooperation and global responsibility with increasing pressures toward competition and strategic positioning (Knight & de Wit, 2018). In this context, higher education institutions must adopt frameworks that prioritize the systematic measurement and analysis of global engagement in order to fully realize its institutional value.

Within this context, measurement plays a constitutive role in shaping institutional reality. What institutions are able to see, compare, and evaluate is determined by the architecture of their data systems. Accordingly, measurement is not simply a technical function of reporting, but a governance mechanism that structures decision-making and prioritization. Major global ranking systems aim to identify top-performing universities, yet they assess only a small fraction of institutions worldwide and rely on selective criteria (Altbach & Hazelkorn, 2018). As a result, rankings function as a zero-sum system that offers limited utility for most institutions. The Johnson Ranking of Global Engagement addresses these limitations by expanding measurement into a comprehensive, institution-specific diagnostic framework. When measurement systems are incomplete or inconsistent, institutional activity may be extensive yet analytically invisible. Conversely, when measurement is coherent and multidimensional, it becomes possible to translate dispersed activity into actionable knowledge.

This study therefore situates measurement as a central analytic lens for understanding global engagement. Rather than treating data as a passive reflection of institutional behavior, it assumes that measurement systems actively construct the boundaries of institutional awareness. In the absence of integrated diagnostic frameworks, global engagement remains partially visible and unevenly interpreted, limiting its strategic utility. Although efforts to refine global ranking methodologies continue, they remain largely externally oriented and insufficient for internal decision-making (Altbach & Hazelkorn, 2018). In contrast, the Johnson Ranking of Global Engagement (JRGE) shifts the focus from comparison to diagnosis by leveraging institutional data to evaluate how global engagement is structured, measured, and managed within a single institution. In doing so, it responds to the growing demand for more meaningful, context-specific evaluation by making global engagement visible, interpretable, and actionable at the institutional level. It is within this gap between activity and visibility that JRGE is positioned as an intervention, functioning as a diagnostic framework that renders distributed global activity legible across institutional domains.

Taken together, these propositions establish the conceptual basis for examining global engagement as a measurable, distributed, and structurally fragmented institutional phenomenon. Over the past several decades, internationalization in higher education has evolved from a values-driven model grounded in cooperation and exchange to one increasingly shaped by competition, commercialization, and strategic positioning. This shift raises critical questions about whether internationalization has fulfilled its intended purposes and what principles should guide its future. As institutions navigate this tension, there is a growing need for more intentional, responsible, and analytically grounded approaches to understanding and managing global engagement (Knight and de Wit, 2018). As internationalization has shifted from a values-based practice to a strategic institutional imperative, the absence of coherent measurement frameworks has become a critical limitation, necessitating tools such as JRGE to render global engagement visible, interpretable, and governable. In this sense, measurement does not merely describe global engagement—it defines the conditions under which it can be governed. This justifies the need for diagnostic approaches that move beyond external reputational metrics toward internally coherent systems of visibility and interpretation.

JRGE Framework

The Johnson Ranking of Global Engagement (JRGE) is a multidimensional, data-driven framework designed to diagnose how global engagement is structured, distributed, and operationalized within higher education institutions. As higher education systems are increasingly shaped by national policy, global pressures, and shifting institutional priorities, internationalization has evolved into a matter of strategic and national importance extending beyond the traditional domains of teaching, research, and service (Damiyano, 2022). Unlike conventional global ranking systems that prioritize external comparison and reputational positioning, JRGE is constructed as an internal analytic instrument. Despite its name, JRGE does not produce ordinal rankings, but instead generates structured diagnostic insights. Its purpose is not to rank institutions against one another, but to render institutional global engagement visible, interpretable, and actionable through the systematic organization and analysis of institutional data.

JRGE is grounded in the premise that global engagement operates as a distributed institutional system spanning multiple domains of activity. To capture this complexity, the framework is organized into four primary domains: Academic, Economic, Social/Cultural, and Governance/Infrastructure. Together, these domains represent the primary mechanisms through which internationalization is enacted and sustained within institutions. At its core, internationalization involves the integration of international and intercultural dimensions into teaching, research, and institutional service. As globalization accelerates, this integration becomes increasingly unavoidable, requiring institutions to align standards while maintaining the quality and integrity of educational provision (Damiyano, 2022). This alignment is critical not only for institutional competitiveness but also for ensuring that students are prepared to operate in globally interconnected environments and that disparities in educational quality are minimized.

The **Academic** domain captures the extent to which global perspectives are embedded within core academic functions. This includes internationally oriented curricula, global research collaborations, and faculty engagement in international scholarship. Beyond presence, this domain highlights the limitations of existing measurement practices, particularly in capturing the depth and impact of research globalization.

The **Economic** domain focuses on institutional engagement through partnerships, funding relationships, and externally oriented collaborations. This includes formal agreements such as memoranda of understanding (MOUs), industry partnerships, and participation in global networks. Critically, JRGE does not assume that the existence of partnerships equates to impact; instead, it evaluates the extent to which these relationships are coordinated, tracked, and strategically aligned.

The **Social and Cultural** domain reflects the experiential and relational dimensions of global engagement. This includes international student enrollment, study abroad participation, intercultural programming, and campus-based global experiences. While often the most visible dimension of internationalization, this domain frequently reveals a reliance on activities that are not fully integrated into institutional strategy.

The **Governance and Infrastructure** domain assesses the organizational capacity required to sustain global engagement. This includes administrative structures, policy frameworks, data systems, and cross-unit coordination mechanisms. This domain is particularly critical, as it often reveals fragmentation, lack of centralized oversight, and underdeveloped systems for managing internationalization.

Within each domain, JRGE operates through a set of dimensions and indicators derived from institutional data sources. These indicators are not intended to produce a single composite score; rather, they function as a diagnostic grid that reveals patterns of concentration, absence, and fragmentation. Central to this approach is the assertion that internationalization must be conceptually bounded and precisely defined. When treated as a catch-all term for anything "foreign," it loses analytical utility and cannot be rigorously operationalized or assessed (Damiyano, 2022).

A defining feature of JRGE is its treatment of missing or incomplete data. Rather than viewing data gaps solely as methodological limitations, the framework interprets them as analytically meaningful signals of institutional structure. Absence of data may indicate fragmentation across units, lack of reporting systems, or insufficient strategic prioritization. In this way, JRGE moves beyond measurement to diagnose the organizational conditions that shape global engagement. The framework also emphasizes the distinction between activity and measurability. Institutions may demonstrate substantial global engagement in practice while lacking the infrastructure to document or evaluate it. Historically, internationalization has often been conceptualized as a collection of activities rather than as an integrated institutional system, resulting in a disconnect between practice and measurement (Damiyano, 2022). JRGE addresses this "lived–measured disjuncture" by aligning institutional activity with corresponding data structures, enabling a transition from anecdotal awareness to systematic analysis.

Finally, JRGE responds directly to a critical limitation in existing internationalization assessment models. Prior approaches have focused primarily on inputs and outputs—such as program counts or participation rates—while neglecting outcomes, impact, and systemic coherence. More robust evaluation requires attention to the internationalization process, its resulting impacts, and the degree to which institutional efforts are strategically aligned (De Wit, 2010; Knight, 2008 qtd. in Damiyano, 2022). This insight is foundational to JRGE. The framework moves beyond activity-based indicators toward a model that captures process, outcomes, and institutional integration as core dimensions of global engagement. By doing so, JRGE provides a mechanism for institutions to identify gaps, prioritize interventions, and align global activity with strategic objectives. Ultimately, JRGE functions as a diagnostic tool for institutional self-assessment. It operationalizes the premise that global engagement is not only a distributed institutional function, but one that must be made visible, measurable, and strategically coordinated in order to be effectively governed in an increasingly borderless higher education landscape.

Methodological Framework

The Johnson Ranking of Global Engagement (JRGE) is operationalized through a structured, multi-phase methodological approach designed to translate the conceptual framework into a systematic institutional analysis. Higher education is increasingly embedded within global, national, and local dynamics, reflected in cross-border mobility, the influence of rankings, and the expansion of technology-driven education models. As a result, internationalization has become a central and enduring component of institutional strategy, despite variation in its definition and implementation (Aziz, Ismail, & Abdullah, 2014). This methodology is grounded in the premise that global engagement is not directly observable as a singular construct, but must be inferred through the aggregation, organization, and interpretation of distributed institutional data. Because internationalization operates across multiple interconnected domains, traditional, narrowly defined metrics are insufficient to capture its complexity. JRGE therefore provides a methodological structure to systematically capture, interpret, and manage global engagement at the institutional level. To systematically capture this distributed activity, institutional data were purposefully gathered from diverse operational modules, including the office of institutional research, international programs, individual academic departments, sponsored programs, research administration, and student affairs enrollment systems. This decentralized data collection methodology intentionally mirrors the structural configuration of the university itself.

Research Design

This study employs a descriptive, diagnostic case study design focused on institutional-level analysis. The purpose is not hypothesis testing, but structured interpretation—specifically, to map how global engagement is manifested, measured, and coordinated within a single higher education institution. As higher education institutions face increasing expectations for accountability, internationalization efforts are subject to greater scrutiny, requiring clear and purposeful evaluation. Assessment must align with institutional motivations: economically driven internationalization may be evaluated at the institutional level, while learning-focused goals, such as intercultural competence, require measurement at the individual level (Coelen, 2019). Accordingly, measurement must be purpose-driven, multi-level, and aligned with institutional intent.

The JRGE framework addresses these measurement gaps by operationalizing global engagement across four distinct structural domains, thereby prioritizing institution-specific data and institutional learning over abstract, reputation-driven metrics. Rather than functioning as a rigid or punitive assessment, the framework operates as an internal diagnostic system designed to capture how universities explicitly navigate global engagement within highly complex organizational ecosystems. By emphasizing qualitative evidence-tracking alongside standardized indicator assessments, the methodology grounds institutional evaluation within the realities of localized administrative operations. The research design is therefore aligned with JRGE's function as an internal analytic instrument rather than a comparative ranking system. It recognizes that global engagement cannot be evaluated through a single metric or level of analysis, instead requiring a multidimensional, context-sensitive approach.

Methodological Approach

The methodological approach integrates both quantitative and qualitative elements. Quantitative analysis is reflected in the aggregation of institutional indicators across domains, while qualitative interpretation is used to assess patterns of coherence, fragmentation, and alignment. This mixed analytic logic reflects the complexity of internationalization as both a measurable and interpretive phenomenon. Existing measurement efforts, including certification frameworks and global rankings, demonstrate the growing emphasis on assessing internationalization. However, these approaches are often limited in scope and insufficient for internal institutional analysis. Effective measurement must therefore be context-sensitive and adaptable to institutional conditions (Coelen, 2019). JRGE addresses this by: 1) Rejecting one-size-fits-all evaluation models in favor of a customizable, institution-specific approach, 2) Moving beyond rankings to focus on internal structure, function, and measurability, and 3) Allowing for contextual adaptation based on institutional realities. This positions JRGE as a contextually grounded alternative to standardized assessment models.

Data Sources and Collection

JRGE relies exclusively on institutional data sources, emphasizing internal validity and contextual specificity. Data are collected from multiple units to reflect the distributed nature of global engagement. To capture the distributed nature of global activities across the institution, empirical data were systematically compiled from a variety of operational positions. Primary administrative data were extracted from the office of institutional research, global engagement units, and international programs to map enrollment trends and formal agreements. These data were further enriched by auditing curriculum profiles within academic colleges and departments, tracking cross-border grant documentation through sponsored programs and research administration, and reviewing domestic student mobility records via enrollment management systems. Gathering data from these disparate administrative siloes allowed the JRGE to directly expose the institutional coordination gaps and structural fragments detailed in the subsequent findings. Data collection is intentionally decentralized, mirroring the structural fragmentation often present in global engagement activities. This approach captures both formalized and informal dimensions of internationalization.

Domains, Dimensions, and Indicators

The JRGE framework is operationalized through a hierarchical structure of domains, dimensions, and indicators: Domains (Academic, Economic, Social/Cultural, Governance/Infrastructure) represent broad areas of institutional activity; Dimensions specify subcategories within each domain; Indicators are observable or measurable data points derived from institutional sources. Indicators are selected based on: Relevance to global engagement; Availability within institutional systems; Interpretive value in revealing patterns of activity or absence. Indicators are not treated as isolated performance metrics; their analytic value emerges through patterns across domains.

Analytic Strategy

Analysis proceeds in three stages:

1. Data Mapping

Institutional data are organized according to the JRGE domain structure, identifying where global engagement activities are documented and how they are distributed.

2. Pattern Identification

Indicators are analyzed collectively to identify patterns of: Concentration (high activity or strong documentation), Absence (limited or missing data), Fragmentation (activity without coordination).

3. Interpretive Diagnosis

Patterns are interpreted in relation to institutional structure and strategy, assessing: alignment across domains, gaps between activity and measurement, and strength of governance and coordination mechanisms. This three-stage process reflects JRGE's shift from descriptive measurement to interpretive institutional diagnosis.

Treatment of Missing Data

A defining feature of JRGE is its treatment of missing or incomplete data as analytically meaningful. Rather than excluding gaps, the framework interprets them as indicators of institutional structure. Missing data may reflect a lack of centralized data systems, decentralized or siloed operations, absence of strategic prioritization, or informal/undocumented activity. By treating absence as signal rather than error, JRGE captures dimensions of institutional functioning typically overlooked in traditional models.

Limitations

Several limitations should be acknowledged. First, reliance on institutional data introduces variability in quality, availability, and consistency. Second, the framework is context-dependent and does not aim for cross-institutional comparability. Third, the interpretive nature of the analysis requires caution to avoid over-attribution of meaning to observed patterns.

Summary

The JRGE methodological framework translates a complex conceptual model into a structured analytic process capable of capturing the distributed and often fragmented nature of global engagement. By integrating data mapping, pattern identification, and interpretive diagnosis, it enables institutions to move beyond surface-level metrics toward a deeper understanding of how internationalization is operationalized and governed. In an environment shaped by learner-centered approaches, diverse forms of engagement, and uneven access to global experiences, this methodology is designed not only to document visible activity, but also to reveal where critical dimensions of global engagement remain unmeasured. In doing so, it reinforces JRGE's central function as a diagnostic tool for institutional self-understanding and strategic action.

Operationalization and Scoring Procedures of the JRGE

The Johnson Ranking of Global Engagement (JRGE) is operationalized as a comprehensive matrix consisting of four primary domains, each containing five specific indicators, yielding 20 total evaluative touchpoints. Across all four domains, the indicator measuring "International Students, Faculty and Staff" serves as a cross-cutting anchor to trace how human capital intersects with institutional infrastructure. Each of the 20 indicators is evaluated using a dual-layered assessment mechanism: a quantitative rubric score and a qualitative rationale validation. The quantitative dimension utilizes a 5-point scale to assess the maturity of institutional data and implementation: *0 represents no evidence; 1 indicates an emerging status; 2 denotes early development; 3 signifies an established and developing framework; 4 indicates a strong, integrated system; and 5 marks a transformative, equity-centered operation.* Concurrently, evaluators audit each indicator against four distinct organizational rationales to document institutional mechanics: *Activity* (programs or documented actions), *Ethos* (institutional commitment and equity orientation), *Competency* (demonstrated capability or effectiveness), and *Process* (implementation, assessment coordination, and continuous improvement mechanisms). These rationales are verified categorically as Yes, No, or Other with space for evaluator notes.

To calculate the final institutional diagnosis, indicator scores are aggregated within their respective domains, producing domain totals out of 25 possible points. These four domain subtotals are then combined to generate a final composite JRGE score out of 100 points. The composite score maps directly onto an automated institutional performance hierarchy: *Developmental (0–25), Emerging (26–50), Established (51–75), Advanced (76–90), and Transformative (91–100).* This multi-tiered calculation shifts the focus of global university evaluation away from external competition toward internal systemic visibility.

Case Application: JRGE Analysis at a Public R1 University in the Southwestern United States

This study applies the Johnson Ranking of Global Engagement (JRGE) framework to a Public R1 University in the Southwestern United States as a single-institution case to demonstrate the framework's diagnostic utility. A public R1 university in the southwestern United States was selected as an information-rich case because of its rapidly evolving internationalization infrastructure, centralized global engagement leadership, and ongoing efforts to strengthen institutional assessment of global engagement. The case institution serves as an appropriate site due to its precipitous institutional growth, explicit emphasis on community engagement, and increasing investment in global initiatives within a complex, decentralized organizational structure. The purpose of this case application is not to evaluate the case institution in normative terms, to operationalize JRGE through systematic mapping of how global engagement is structured, distributed, and measured across the institution. In doing so, the analysis demonstrates how JRGE functions as an internal analytic instrument capable of identifying patterns of concentration, fragmentation, and institutional alignment.

Institutional Context

The case institution is a large, urban, public research university located in the southwestern United States, characterized by a highly diverse student population and a strong institutional emphasis on access, workforce development, and community impact. As part of its broader strategic trajectory, the institution has increasingly positioned itself as a globally engaged institution, reflected in expanded international partnerships, growth in international student enrollment, and participation in global research networks. At the same time, although the institution has recently centralized multiple global engagement functions under a unified institutional structure, internationalization activities remain distributed across academic colleges, research offices, student services, and administrative units. This reflects a broader institutional reality in which global engagement operates across the entire university ecosystem, making comprehensive assessment and coordination inherently complex. This structural configuration provides a critical context for examining the “lived–measured disjuncture” that JRGE is designed to capture—where institutional activity exceeds the capacity of centralized systems to document and interpret it.

Data Mapping Across JRGE Domains

Data were collected from institutional sources and systematically organized according to the four JRGE domains: Academic, Economic, Social/Cultural, and Governance/Infrastructure. Consistent with the methodological framework, this stage emphasizes data mapping rather than evaluation, identifying where global engagement is documented and how it is distributed across institutional units. The mapping process revealed variation not only in the presence of global engagement activities, but in the degree to which those activities are systematically captured, coordinated, and made visible.

Within the **Academic** domain, global engagement is evidenced through internationally oriented curricula, faculty-led study abroad programs, and participation in collaborative research initiatives. However, data collection mechanisms related to research globalization are distributed across semi-autonomous institutional units. While distinct capabilities to track international co-authorship, cross-border grant collaborations, and global knowledge networks reside within a department of research, individual academic colleges, and the institution’s libraries, these metrics have historically remained unaggregated. Consequently, significant global research activity exists, though its full scope has not been consistently visible through unified institutional reporting systems. This indicates the presence of activity without corresponding institutional visibility. Although missing data is an analytical signal within the JRGE, the data is pullable; it simply lacked an integrated catalyst to synthesize it cohesively, which has been somewhat alleviated through their office of globalization’s most recent annual report.

In the **Economic** domain, the institution demonstrates a strong presence of formal partnerships, including memoranda of understanding (MOUs) and participation in external networks. However, these partnerships are primarily documented as static agreements, with limited mechanisms for assessing their functional impact, utilization, or alignment with institutional strategy.

The **Social and Cultural** domain represents the most visible and consistently documented area of global engagement. This includes international student enrollment, study abroad participation, and campus-based intercultural programming. These activities are more systematically tracked, reflecting both their visibility and their alignment with institutional reporting priorities. At the same time, they are not consistently integrated into broader institutional strategy or cross-domain assessment structures.

The **Governance and Infrastructure** domain reveals the greatest degree of variation and limitation. Dedicated administrative units and centralized global offices exist to support internationalization, although institutional data systems and reporting mechanisms remain in varying stages of integration across units. Recent efforts, including annual reporting initiatives and expanded collaboration with sponsored programs, demonstrate ongoing progress toward more coordinated institutional visibility and assessment.

Patterns of Institutional Structure

Building on the data mapping stage, the JRGE framework enables identification of cross-domain patterns that reflect underlying institutional structure:

First, concentration of visibility is evident in domains where activities are highly visible and easily quantifiable, particularly within the Social and Cultural domain. These areas benefit from established reporting systems and institutional emphasis, resulting in more complete and accessible data.

Second, fragmentation across domains is evident in both the Academic and Economic domains, where substantial activity exists but is inconsistently documented, coordinated, or integrated. This fragmentation reflects decentralized ownership of global engagement and the absence of unified data infrastructure.

Third, misalignment between activity and measurement emerges as a defining institutional characteristic. While the institution demonstrates extensive global engagement in practice, the systems required to capture, evaluate, and strategically align this activity are unevenly developed.

Data Absence as Institutional Signal

Consistent with the JRGE methodological approach, areas of missing or incomplete data are interpreted as analytically meaningful. In this case, data gaps are most pronounced within the Governance/Infrastructure domain and in aspects of academic research globalization. These absences suggest not a lack of activity, but a lack of centralized reporting, coordination, or strategic prioritization. For example, the absence of comprehensive data on international research collaboration indicates that such activity likely occurs at the faculty level without institutional aggregation or visibility. In this sense, data absence functions as an indicator of institutional structure, revealing where global engagement exists but remains unmeasured.

Empirical Diagnostic Results: The Case Institution Application

Application of the JRGE scoring instrument to the institutional data at the case institution yielded a composite score of 54 out of 100, placing the university within the *Established with Developing Integration* tier of internationalization maturity. While core student mobility metrics and localized administrative structures are highly mature, significant variances emerge across academic tracking systems and decentralized governance nodes. As illustrated in Table 1, the institutional metrics reflect the empirical domain-level results, indicator scores, and corresponding strategic signals.

Table 1: JRGE Empirical Diagnostic Matrix for Institutional Application

JRGE Domain	Evaluated Indicator	Score (0–5)	Domain Subtotal	Institutional Performance Tier	Diagnostic Status & Actionable Signal
Social and Cultural	1. International Students, Faculty & Staff	4			
Social and Cultural	2. Global Consciousness	3			
Social and Cultural	3. Global Programs/Activities	4			
Social and Cultural	4. Inter-culturation Integration	3			
Social and Cultural	5. Student Exchange/Study Abroad	4	**18 / 25**	**Established**	High reporting maturity and structured documentation; activity operates independently of central strategic planning.
Internal Business & Economic	1. International Students, Faculty & Staff	4			
Internal Business & Economic	2. Brand Image/Recognition	3			
Internal Business & Economic	3. International Partnerships	2			
Internal Business & Economic	4. Funding/Fundraising for Global Activities	2			
Internal Business & Economic	5. Joint Ventures/Degree Programs	1	**12 / 25**	**Emerging**	Static MOU documentation is robust, but lacks continuous operational tracking or active evaluation.
Academic & Teaching/Learning	1. International Students, Faculty & Staff	4			
Academic & Teaching/Learning	2. International Research Collaboration	2			

JRGE Domain	Evaluated Indicator	Score (0–5)	Domain Subtotal	Institutional Performance Tier	Diagnostic Status & Actionable Signal
Academic & Teaching/Learning	3. Worldwide Publications/Citations	2			
Academic & Teaching/Learning	4. Academic Cross-Border Engagement	2			
Academic & Teaching/Learning	5. Visiting Scholars/Faculty-Staff Mobility	1	**11 / 25**	**Emerging**	Significant distributed activity exists across units; requires a centralized mechanism to synthesize system visibility.
Governance: Leadership & Admin	1. International Students, Faculty & Staff	4			
Governance: Leadership & Admin	2. Administrative Support/Involvement	3			
Governance: Leadership & Admin	3. International Offices	3			
Governance: Leadership & Admin	4. Closely-Coupled Systems	1			
Governance: Leadership & Admin	5. Awards/International Recognition	2	**13 / 25**	**Emerging**	Centralized leadership structure is established, but deeply entrenched data silos across administrative nodes persist.
OVERALL MATRIX CONCLUSION	**COMPOSITE JRGE SCORE**		**54 / 100**	**Established with Developing Integration**	**Final Institutional Performance Tier Evaluation Summary**

Note: Scores compiled using the JRGE Digital Evaluation Tool based on the case institution’s internal institutional data, qualitative inquiry and division reports.

Summary

The JRGE application at the case institution demonstrates the framework's capacity to move beyond surface-level indicators of internationalization toward a more structured understanding of institutional organization. By integrating data mapping with pattern identification, the analysis reveals a system characterized by concentrated visibility, underlying fragmentation, and misalignment between institutional activity and measurement capacity. This case application establishes the empirical foundation for the subsequent findings and discussion by illustrating how JRGE translates distributed institutional data into diagnostically meaningful insights about global engagement.

Findings

The application of the Johnson Ranking of Global Engagement (JRGE) framework to the case institution reveals a set of consistent and analytically significant patterns that extend beyond descriptive observation. These findings reflect not only where global engagement is present, but how it is structured, measured, and governed within the institution. Five primary findings emerge from the analysis:

1. Global Engagement is Highly Visible but Unevenly Distributed

The analysis indicates that global engagement at the university is both extensive and unevenly distributed across institutional domains. The Social and Cultural domain—including international student enrollment, study abroad participation, and intercultural programming represent the most visible and systematically documented area of activity. These forms of engagement are supported by established reporting systems and closely aligned with institutional priorities related to access, diversity, and student experience. In contrast, global engagement within the Academic and Economic domains is less consistently documented, despite evidence of substantial activity. Internationalized curricula, faculty-led initiatives, and collaborative research are present, but their scope and impact are not systematically captured at the institutional level. This uneven distribution suggests that visibility is shaped less by the presence of activity and more by the existence of measurement infrastructure.

2. Partnership Expansion Occurs Without Systematic Evaluation

The findings reveal a significant expansion of formal international partnerships, particularly within the Economic domain. The university maintains numerous memoranda of understanding (MOUs) and participates in a range of global networks. However, these partnerships are primarily documented as formal agreements rather than as active, measurable engagements. Although formalized partnership assessment mechanisms have historically been limited, the institution has recently implemented new partnership evaluation procedures designed to align agreements with institutional strategic priorities and assess partnership activity over time: alignment with institutional improvement strategies and strategic plans, mid-agreement check-ins, activity reporting, the functional use of partnerships, their alignment with institutional strategy, and their contribution to academic, research, or student outcomes. Historically, partnerships functioned primarily as indicators of global presence rather than components of a coordinated institutional system. The recently implemented partnership review process represents an important shift toward strategic evaluation and continuous institutional oversight.

3. Research Globalization Visibility is Constrained by Distributed Tracking Infrastructure

The empirical application of the JRGE Academic domain underscores that while extensive global research activity occurs across the university, institutional visibility remains constrained by a distributed data tracking infrastructure. Highly sophisticated data gathering capabilities reside natively within separate entities, including the research department and within sponsored programs, individual academic colleges, and university's libraries. These units regularly track indicators such as international co-authorship, cross-border grant funding, and global publication impact on an ad-hoc or unit-specific basis. Crucially, the diagnostic score of 11/25 in this domain does not signal an institutional failure or an absence of tracking activity; rather, it reflects a historical lack of centralized aggregation and reporting. Until recent institutional annual reporting mandates established a collaborative motivation to synthesize these metrics cohesively, there was minimal systemic integration across these units. This structural arrangement generates an artificial disjuncture between the extensive international research actually produced by faculty and its collective visibility on centralized institutional reporting platforms. This finding highlights a clear opportunity for a newly consolidated department for global engagement to act as a collaborative hub, working across capable administrative nodes to bring unified visibility to the case university's international research footprint.

4. Governance and Data Systems Reflect Transitional Integration

A significant finding emerging from the analysis is that fragmentation in governance and data infrastructure can persist even within increasingly centralized internationalization systems. While the institution has recently consolidated multiple global engagement offices under a unified institutional structure, global engagement activities continue to span numerous academic and administrative units across the university. This fragmentation is reflected in decentralized data ownership across units, inconsistent reporting practices, and developing cross-domain integration mechanisms. As a result, global engagement functions as a distributed system without centralized visibility or coordination. This limits the institution's ability to assess overall performance, identify gaps or redundancies, and align global activity with strategic priorities.

5. Misalignment Between Activity and Measurement Defines Institutional Reality

Taken together, the findings point to a defining institutional condition: ongoing tension between institutional activity and institutional measurement capacity. The institution demonstrates substantial and diverse global activity across domains, yet lacks the integrated measurement systems necessary to capture and interpret that activity coherently. This misalignment produces a "lived–measured disjuncture," in which institutional practices exceed measurement capacity, visibility is uneven across domains, and decision-making is constrained by incomplete information. Within this context, measurement is not simply a technical limitation but a structural condition that shapes what can be known, managed, and governed. Collectively, these findings demonstrate that institutional visibility is not a function of activity itself, but of the systems through which activity is measured, aggregated, and interpreted.

Summary of Findings

The JRGE analysis reveals that global engagement at the case university is not absent, but unevenly visible, inconsistently measured, and structurally fragmented. The institution exhibits strong engagement in highly visible domains, expanding partnership activity without systematic evaluation, limited measurement of research globalization, and fragmented governance systems. However, the analysis also indicates that the university has begun implementing centralized assessment and partnership evaluation mechanisms intended to address these gaps incrementally over time. Collectively, these findings demonstrate that the primary challenge is not the presence of global engagement, but the absence of comprehensive integrated measurement systems capable of rendering that engagement visible and actionable. More broadly, these patterns reflect a wider shift in higher education, where internationalization has evolved from a mission of collaborative capacity building into a high-stakes pursuit of global rankings and revenue, introducing ethical risks and reinforcing competitive, status-driven behavior (Knight, 2013, p. 84). Continued refinement of evaluation procedures, centralized reporting, and assessment priorities should progressively reduce these gaps as the institution's internationalization infrastructure continues to mature.

In this context, global engagement expands in scope without corresponding institutional oversight. The case institution illustrates how this disconnect produces a gap between visible global presence and the institutional capacity required to manage it effectively. Bridging this gap requires not only increased activity, but the development of integrated, diagnostic measurement systems capable of aligning internationalization with institutional strategy, accountability, and core academic values.

Discussion

This study examined how global engagement is structured, measured, and governed within a contemporary higher education institution through the application of the Johnson Ranking of Global Engagement (JRGE). The findings from the case institution reinforce a central premise: global engagement is not absent within institutions, but unevenly visible, structurally fragmented, and insufficiently measured. As such, the primary challenge is not the expansion of activity, but the development of systems capable of rendering that activity coherent, interpretable, and actionable. As higher education continues to evolve within a global regime increasingly shaped by quality assurance, accreditation, and standardization, the need for structured diagnostic tools to make institutional activity visible becomes more urgent (Zapp & Ramirez, 2019, p. 16).

Global Engagement as a Distributed Institutional Function

The findings support the conceptualization of global engagement as a distributed institutional function rather than a centralized domain. At the case university, global activity is embedded across academic, economic, social, and administrative domains, each operating with varying degrees of autonomy and coordination. This distribution enables breadth but produces fragmentation in institutional knowledge. This pattern aligns with broader shifts in internationalization, particularly the move from mobility-centered models toward more inclusive “internationalization at home” approaches that engage the full academic community while emphasizing sustainability, societal impact, and global–local integration (de Wit & Altbach, 2020, p.17). The concentration of visibility in student-facing, mobility-driven domains at the case institution illustrates the limitations of earlier models and reinforces the need to reconceptualize global engagement as system-wide rather than program-specific.

Accordingly, institutions can no longer rely on visible mobility indicators as proxies for global engagement. Instead, they must adopt diagnostic frameworks, such as JRGE, capable of capturing embedded, less visible, and institutionally distributed forms of internationalization. In this sense, JRGE operationalizes this global shift by making dispersed engagement measurable and strategically actionable. This case study further demonstrates that centralization alone does not eliminate fragmentation. Even within institutions that consolidate global functions administratively, internationalization remains embedded across decentralized academic and operational units. This suggests that the challenge facing institutions is not simply organizational structure, but the creation of integrated systems capable of coordinating distributed forms of engagement.

Measurement as a Condition of Governance

A central contribution of this study is the demonstration that measurement is not merely technical, but constitutive of institutional governance. What institutions are able to see, prioritize, and manage is structurally shaped by their data systems. In the absence of integrated measurement frameworks, global engagement remains partially visible and unevenly interpreted. The case study illustrates how this constrains decision-making. At the same time, the institution demonstrates that measurement systems evolve incrementally through administrative coordination, annual reporting development, and cross-unit collaboration. Domains with established reporting systems, particularly Social and Cultural, gain visibility and institutional priority, while areas such as research globalization remain underrepresented despite significant activity. This reinforces the argument that data systems do not simply reflect reality; they actively construct it. Moreover, as prior research emphasizes, meaningful assessment must move beyond participation metrics to evaluate outcomes, including changes in student learning, values, and intercultural competence (Ching et al., 2012, p. 211). Improving global engagement, therefore, requires not only expanding activity, but strengthening the systems through which that activity is documented, interpreted, and governed.

From Activity to Alignment

The findings reveal a persistent misalignment between global engagement as practiced and as measured. Institutions may exhibit extensive activity while lacking the capacity to assess coherence, impact, or strategic alignment. This is particularly evident in the expansion of partnerships without evaluation mechanisms and in the limited visibility of research globalization.

Addressing this misalignment requires a shift from activity accumulation to strategic alignment, including: linking partnerships to institutional priorities, integrating engagement across units, assessing outcomes rather than participation. In this context, internationalization must move beyond accumulation toward coordination and intentionality. As global flows of people, ideas, and technology reshape higher education, internationalization reflects a dynamic interplay of socio-economic, political, and cultural forces (Frezghi & Tsegay, 2019, p. 654), further underscoring the need for structured, institutionally coherent approaches.

Implications for Practice

These diagnostic findings carry significant implications for the evolution of institutional governance and strategic planning. First, the application demonstrates that universities must actively transition from loosely-coupled operational silos toward closely-coupled data systems to ensure that international strategic initiatives are supported by real-time operational metrics. Second, the framework illustrates that evaluating data availability serves as a valuable proxy for tracing internal power dynamics, resource distribution, and operational priorities. Finally, the JRGE highlights that institutions must move away from prestige-driven external rankings that ignore localized contexts, shifting instead toward equity-centered internal assessments that promote institutional learning and continuous capacity building.

Methodological Limitations and Framework Transferability

While this study offers rich insights into the mechanics of institutional data tracking, several limitations must be acknowledged. First, because the JRGE framework was applied here as a single institutional case study, the specific scoring dynamics and structural gaps identified are contextualized within a large, public, urban research university designated as a minority-serving institution. Consequently, the literal baseline scores are not directly generalizable to smaller private colleges, community colleges, or highly centralized international university systems. Furthermore, the reliance on an internal evaluator model introduces inherent subjectivity during the qualitative rationale auditing process (evaluating Activity, Ethos, Competency, and Process), which may limit immediate external standardization.

To address these constraints, future research must focus on the empirical transferability and validation of the JRGE model. Future investigators should apply this 20-indicator matrix across diverse institutional archetypes, including land-grant universities, liberal arts colleges, and international institutions, to evaluate how variations in organizational size and funding impact domain scoring. Cross-institutional benchmarking studies would allow researchers to statistically validate the indicators and refine the rubric definitions. Additionally, creating external peer-review auditing procedures for the qualitative rationales would mitigate internal evaluator bias, transforming the JRGE into a validated, scalable, and highly reliable instrument for comparative higher education research.

JRGE as a Diagnostic Framework

This study positions JRGE as a scalable diagnostic framework for institutional self-assessment. Unlike traditional rankings, which emphasize external comparison, JRGE generates internally relevant insights by organizing institutional data into meaningful patterns.

Its primary contribution lies in its ability to: 1) translate distributed activity into structured knowledge, 2) identify gaps between activity and measurement, and 3) interpret absence as an indicator of institutional structure. In this sense, JRGE functions not only as a diagnostic instrument, but also as a developmental planning framework capable of helping institutions identify assessment gaps, prioritize strategic initiatives, and monitor institutional progress over time. In doing so, JRGE shifts the focus from comparison to diagnosis, enabling institutions to understand not how they rank globally, but how they function internally.

Conclusion

In an era where internationalization has evolved from collaborative capacity building to a competitive, status-driven enterprise, institutions face increasing pressure to demonstrate both global presence and measurable impact. However, as this study demonstrates, expansion without corresponding measurement and governance systems produces a condition in which activity exceeds institutional understanding. The case also demonstrates that institutional progress toward integrated internationalization assessment is iterative rather than immediate. Limiting the flow of people and ideas constrains innovation and knowledge production, while the absence of clear, governance-oriented policy frameworks results in uneven global engagement and institutional ambiguity (Rajkhowa, 2017, p. 33). This case study illustrates this condition: global engagement is widespread and increasingly coordinated, yet still unevenly visible across institutional domains as measurement systems continue to mature.

This is not an isolated case, but rather reflects broader structural dynamics in higher education. Addressing this challenge requires a fundamental shift from viewing internationalization as a collection of activities to understanding it as a system requiring measurement, coordination, and governance. The JRGE framework offers one pathway forward by providing a structured, data-driven approach to making global engagement visible and actionable. In doing so, it supports a transition from uncoordinated participation in global flows to intentional, strategically aligned, and institutionally governed engagement. Emerging governance structures, partnership review systems, and collaborative reporting mechanisms suggest that diagnostic frameworks such as JRGE may serve not only to identify institutional gaps, but also to support long-term strategic development and organizational learning. Ultimately, the future of internationalization will depend not only on what institutions do globally, but on their capacity to understand, measure, and manage those activities in a coherent and integrated manner. In this sense, this study contributes to a broader reconceptualization of internationalization, not as a set of activities or outcomes, but as a measurable, governable institutional system whose effectiveness depends on the visibility and integration of its underlying structures. Beyond its application at the case institution, the Johnson Ranking of Global Engagement provides higher education institutions with a scalable framework for diagnosing internationalization, identifying institutional priorities, and supporting evidence-informed decision-making across diverse organizational contexts.

References

Abd Aziz, Mohd Ismail, & Doria Abdullah. "Finding the next 'Wave' in Internationalisation of Higher Education: Focus on Malaysia." *Asia Pacific Education Review* [Dordrecht], vol. *15*, no. 3, September 2014, pp. 493–502, https://doi.org/10.1007/s12564-014-9336-7

Altbach, Philip G, & Hans de Wit. "Are We Facing a Fundamental Challenge to Higher Education Internationalization?" *International Higher Education*, no. 93, March 2018, pp. 2–4, https://doi.org/10.6017/ihe.0.93.10414

Altbach, Philip G, & Ellen Hazelkorn. "Measuring Education Quality in Global Rankings:

What's the Likelihood?" *International Higher Education*, no. 95, September 2018, pp. 12–14, https://doi.org/10.6017/ihe.2018.95.10721

Barbosa, Maria Lígia de Oliveira, & Clarissa Eckert Baeta Neves. "Internationalization of Higher Education: Institutions and Knowledge Diplomacy." *Sociologias*, vol. *22*, no. 54, August 2020, pp. 22–44, https://doi.org/10.1590/15174522-104425

Ching, G. S., Palmer, J. D., Cho, Y. H., & Roberts, A. (Eds.). (2012). *The internationalization of East Asian higher education: Globalization's impact*. Palgrave Macmillan. https://doi.org/10.1057/9781137002006

Coelen, R. J. (2019). *Changing landscapes of internationalisation of higher education: The needs of associate degree programs*. NHL Stenden University of Applied Sciences; Chain5. https://www.chain5.net/wp-content/uploads/2013/12/Reader-IoHE-Ad-Programs-nhl-stenden.pdf

Damiyano, D., 2022, 'Internationalisation of higher education: A conceptual framework', in N. Dorasamy & G. Mugano (eds.), *Perspectives on comprehensive internationalisation of higher education*, AOSIS Books, Cape Town, pp. 1–27. https://doi.org/10.4102/aosis.2022.BK297.01

Frezghi, Tesfaldet Ghilay, & Samson Maekele Tsegay. "Internationalisation of Higher Education in China: A Critical Analysis." *Social Change (New Delhi)* [New Delhi, India], vol. *49*, no. 4, December 2019, pp. 643–58, https://doi.org/10.1177/0049085719886693

Godwin, K. A., & de Wit, H. (Eds.). (2019). *Intelligent internationalization: The shape of things to come.* Brill.

Knight, Jane. "The Changing Landscape of Higher Education Internationalisation - for Better or Worse?" *Perspectives (Association of University Administrators (U.K.))* [London], vol. *17*, no. 3, July 2013, pp. 84–90, https://doi.org/10.1080/13603108.2012.753957

Knight, Jane, & Hans de Wit. "Internationalization of Higher Education: Past and Future." *International Higher Education*, no. 95, September 2018, pp. 2–4, https://doi.org/10.6017/ihe.2018.95.10715

Ledger, Susan, & Colleen Kawalilak. "Conscientious Internationalisation in Higher Education: Contextual Complexities and Comparative Tensions." *Asia Pacific Education Review* [Dordrecht], vol. *21*, no. 4, December 2020, pp. 653–65, https://doi.org/10.1007/s12564-020-09650-0

Mugano, G. & Dorasamy, N., 2022, 'Background', in N. Dorasamy & G. Mugano (eds.), *Perspectives on comprehensive internationalisation of higher education*, AOSIS Books, Cape Town, pp. xxv–xxvii. https://doi.org/10.4102/aosis.2022.BK297.00

Rajkhowa, Gautam. “Internationalising Indian Higher Education: Opportunities, Challenges and the Way Forward.” *Space and Culture, India*, vol. *5*, no. 1, June 2017, pp. 25–35, https://doi.org/10.20896/saci.v5i1.259

Rajkhowa, G. (2024). Internationalisation for all: Rethinking university internationalisation. *Space and Culture, India*, *11*(4), 5–19. https://doi.org/10.20896/saci.v11i4.1341

van Gaalen, Adinda, et al. “National Policies on Education Abroad Outcomes: Addressing Undesired Consequences.” *Education Abroad*, edited by Bernhard Streitwieser et al., 1st ed., Routledge, 2021, pp. 203–17, https://doi.org/10.4324/9780429431463-13

Zapp, M., & Ramirez, F. O. (2019). Beyond internationalisation and isomorphism – the construction of a global higher education regime, *Comparative Education*, https://doi.org/10.1080/03050068.2019.1638103

White-Collar Crime as the Manipulation of Meaning:

Toward an Operative Criminology of Elite Deviance

Eric Cheney

Gary Myers

Department of Sociology, Central Washington University

Abstract

White-collar crime remains conceptually unsettled within criminology despite decades of scholarship. Traditional frameworks emphasize offender status, breach of trust, statutory violation, or organizational culture, yet they often leave under-specified the mechanism through which harmful practices become stabilized within institutions that continue to appear legitimate. This article reconceptualizes white-collar crime as social harm produced through the manipulation of meaning under the cover of legitimate institutions. Drawing on a systems-oriented framework, we develop seven operative forms—distinction, indication, recursion, re-entry, identification, crossing, and void—that describe recurrent institutional operations through which distorted representations become durable. Through analysis of the Enron collapse, supplemented by briefer consideration of the Madoff fraud and Watergate, we demonstrate how these forms converge to sustain large-scale harm within environments saturated with formal compliance and oversight. We introduce the concept of operative dignity to designate the structural integrity of representational systems upon which institutional participation depends. A meaning model is more explanatory than trust-, occupation-, or status-based definitions because it specifies the communicative mechanisms by which harm is produced, concealed, and normalized. White-collar crime is not merely elite deviance; it is a communicative process through which institutions organize harm while preserving legitimacy. One of its deepest powers is not only that it causes harm, but that it manages suspicion. White-collar crime often uses order itself as camouflage.

Key Words: white-collar crime, meaning, operative forms, institutional legitimacy, operative dignity

Introduction: The Problem of How, Not Who

White-collar crime remains one of the most consequential and conceptually unsettled domains in criminology. Standard approaches emphasize trust, occupation, status, or legal category. These tell us who offends—but not how white-collar harm becomes socially possible. The deeper sociological question is how legitimacy, trust, and institutional normalcy help conceal harm. White-collar crime is hard to see because it often appears inside normal procedure.

Edwin Sutherland's intervention remains foundational. By defining white-collar crime as crime committed by a person of respectability and high social status in the course of occupation (Sutherland, 1940, 1949), he forced criminology to confront elite wrongdoing. Yet status is not itself a mechanism. High social position correlates with access to resources and insulation from scrutiny; it does not explain how harmful practices are organized, justified, and rendered durable within institutional systems.

Trust-based approaches advance the conversation. Susan Shapiro's argument to "collar the crime, not the criminal" reframed white-collar crime as the abuse of trust embedded in fiduciary and professional roles (Shapiro, 1990). David Friedrichs deepened this insight: elite offenders occupy positions in which their representations carry epistemic weight, and their authority shapes what counts as credible information (Friedrichs, 2010). Even so, trust-based frameworks often stop at the level of relational injury. They mark the breach but leave the underlying mechanism under-specified. Trust is not merely misplaced confidence; it is an effect of stabilized communicative form.

Legal positivist definitions introduce a further layer of tension. Paul Tappan (1947) insisted that crime must be defined strictly as violation of criminal law. Yet criminological analysis cannot treat statutory definitions as analytically exhaustive. Robert Agnew's formulation of crime as "blameworthy harm" provides a bridge beyond strict positivism (Agnew, 2011), opening conceptual space for injuries that may be legally insulated but socially destructive. White-collar crime frequently occupies precisely this liminal zone, where enforcement asymmetry is not incidental but structural (Black, 1976).

These literatures converge on a persistent observation: white-collar crime is difficult to detect, difficult to prosecute, and unevenly punished. But they do not fully explain why.

We suggest that the difficulty lies at the level of meaning-production. Elite deviance frequently stabilizes not because rules are absent or trust is naïvely given, but because institutional systems themselves structure the conditions under which harm becomes visible, nameable, and actionable. Trust-, occupation-, and status-based definitions describe the actor. A meaning model explains the mechanism.

This reframing does not displace existing criminological theories—strain, opportunity, organizational culture, regulatory enforcement—but specifies the communicative architecture through which their dynamics are mediated and stabilized.

To render this mechanism analytically tractable, we develop seven operative forms of communication—distinction, indication, recursion, re-entry, identification, crossing, and void—that specify recurrent operations through which institutional systems stabilize distorted representations while preserving formal legitimacy. We also introduce operative dignity as a normative standard grounded in the integrity of meaning-making conditions necessary for institutional participation.

Meaning as Institutional Mechanism

Criminological theory has long relied upon implicit theories of meaning. When we analyze deception, fraud, rationalization, neutralization, labeling, or legitimacy, we presuppose an account of how social reality is constructed, stabilized, and contested. In the context of white-collar crime, this omission becomes consequential, because elite deviance frequently operates at the level where meaning itself is structured.

Institutional systems do not simply transmit information; they structure perception. Financial reporting frameworks classify assets, liabilities, revenues, and risks. Audit certifications indicate compliance. Credit ratings summarize complex exposure into ordinal categories. Regulatory classifications define what counts as violation, oversight, or permissible discretion. These are not neutral containers of information. They are structured distinctions that reduce complexity and coordinate expectations.

Systems-theoretic sociology has emphasized that modern society operates through differentiated institutional domains—law, economy, politics, science—each organized around distinct communicative codes and internal criteria of validity (Luhmann, 1995, 2004). Whether or not one adopts this framework in full, it clarifies a crucial point: institutional meaning is structured. It is stabilized through routines, classifications, and decision premises. White-collar crime often intervenes precisely at this level, reshaping classification systems, exploiting definitional flexibility, capturing credibility signals, and fragmenting oversight channels.

This distinction between interpretive and operative models of meaning is analytically important. An interpretive model asks: What did actors believe? How did they rationalize? How did victims misperceive risk? An operative model asks: What distinctions structured the field? Which signals marked credibility? How were reporting cycles reproduced? How was information fragmented? Both levels matter. But without specifying operative structure, explanations risk collapsing into psychological or moral narratives.

This has a direct methodological implication. An operative account does not require that individual actors be dishonest, morally deficient, or psychologically driven by greed. The framework is agnostic about intention. What it asks instead is: what distinctions were operative? What signals were in place? How were reporting cycles structured? How was critique processed? These are questions about institutional architecture, not character. A CFO may have genuinely believed that SPE structures were permissible; an auditor may have been institutionally prevented from seeing the full configuration of risk. The operative framework does not exonerate such actors, but it insists that individual culpability is analytically insufficient as an explanation of how large-scale harm becomes durable. Moral failure may be present—it often is—but it is not the mechanism. The mechanism is communicative and structural, and it can operate through actors who consider themselves to be acting in good faith, precisely because the conditions of legitimate representation have already been compromised at the level of institutional form.

White-collar crime frequently operates in zones of classification ambiguity. Harmful practices may be redescribed as aggressive accounting, creative compliance, or innovative financial engineering. Such redescriptions are not merely rhetorical; they are embedded in institutional documentation, expert testimony, and professional standards. When these descriptions become stabilized, they shape what regulators, courts, and publics can perceive as actionable. This is why trust alone cannot explain durability. Investors trust because distinctions appear stable, signals appear credible, identities appear consistent, and absences appear insignificant. When these conditions are engineered, trust becomes structurally plausible.

To make this argument analytically usable, we specify seven operative forms through which institutional systems stabilize meaning. These forms are not abstract metaphors; they correspond to observable institutional practices. Distinction corresponds to classification and boundary-setting. Indication corresponds to credibility signaling. Recursion corresponds to temporal reproduction of expectations. Re-entry corresponds to reflexive recoding of critique. Identification corresponds to stabilized institutional self-description. Crossing corresponds to movement across regulatory and organizational domains. Void corresponds to managed invisibility and fragmentation.

The Seven Operative Forms: A Grammar of Elite Deviance

We now specify seven operative forms through which institutional systems stabilize distorted representation. These forms are analytic descriptors of recurrent communicative operations observable in organizational and regulatory contexts. White-collar crime rarely depends upon a single operation. Rather, it emerges through configurations of these forms. The grammar we develop is intended to make those configurations visible and comparable across cases.

Street Crime vs. White-Collar Crime: Distinction

Every institutional system operates through distinctions. Law distinguishes legal from illegal; accounting distinguishes assets from liabilities; regulatory regimes distinguish compliance from violation. These boundaries structure perception and determine what can be observed as harmful.

Street crime often violates order openly. White-collar crime often uses the appearance of order as camouflage. Operative form: Distinction.

Distinction becomes criminologically significant when boundaries are strategically manipulated. Elite actors frequently operate in zones of classification flexibility—areas where definitional thresholds are technical, complex, or politically contested. Regulatory arbitrage, aggressive accounting, structured finance, and jurisdictional relocation all illustrate boundary work in practice. Clinard and Yeager's (1980) analysis of corporate crime demonstrated that large organizations often operate at the edge of compliance, using technical interpretation to maintain formal legality while generating harmful outcomes. When harmful practices are relocated into definitional gray zones, criminality becomes difficult to assert without revisiting the boundary itself. Distinction therefore operates as the foundational mechanism of elite deviance: it reconfigures the categories through which wrongdoing would otherwise be recognized.

Harm Carried Through Legitimacy: Indication

If distinction draws boundaries, indication marks one side of the boundary as authoritative. The key issue is not simply offender status. The indicated point is that harm is carried through titles, procedure, expertise, compliance language, and organizational normalcy. Operative form: Indication.

Audit certifications, credit ratings, professional credentials, regulatory approvals, and media endorsements function as signals that reduce uncertainty for external observers. White-collar crime frequently depends on the capture or manipulation of these signals. Investors and regulators rarely have direct access to underlying transactions; they rely on indications. When indications remain aligned, scrutiny diminishes.

John Coffee (2005) has emphasized the structural ambivalence of gatekeepers such as auditors and rating agencies, who are simultaneously monitors and service providers. Friedrichs' concept of "trusted criminals" underscores the cultural dimension of indication: elite offenders often occupy positions that confer epistemic authority, and their representations are presumed credible because of institutional status (Friedrichs, 2010). So long as high-status signals remain intact, isolated doubts struggle to gain traction.

Organizational Reproduction of Meaning: Recursion

White-collar crime frequently unfolds over time rather than as a single discrete act. Organizations recursively reproduce the meanings that make harmful conduct appear ordinary—through internal memos, reporting categories, managerial talk, and technical routines. Operative form: Recursion.

Diane Vaughan's (1996) concept of the normalization of deviance provides a foundational example: repeated rule-bending adjustments can gradually redefine acceptable risk within organizational culture. In financial contexts, quarterly earnings expectations illustrate recursive pressure. Once an organization publicly commits to performance targets, future reporting cycles are structured by those commitments. Recursive reproduction of indicated performance may incentivize increasingly aggressive boundary manipulation. Recursion also generates path dependency: each iteration of distortion becomes the baseline for the next. This helps explain why some white-collar crimes intensify rather than self-correct.

Illegitimacy Inside Legitimacy: Re-entry

White-collar crime does not simply hide from legitimacy; it often hides behind legitimacy. Illegitimacy re-enters the field of legitimacy and operates through it. Operative form: Re-entry.

Re-entry refers to reflexive operations through which systems process critique. Legal systems review legality; corporations conduct internal investigations; regulatory agencies issue interpretive guidance. These reflexive processes can function as accountability mechanisms—or as containment mechanisms. White-collar crime often persists when critique is absorbed procedurally without altering underlying architecture. Internal compliance reviews may ratify contested practices. Legal consultation may reinterpret borderline behavior as permissible. Settlement agreements may resolve violations without structural reform. Braithwaite's (1985) analysis of regulatory negotiation shows how enforcement often proceeds through dialogue rather than punishment, which can enable organizations to recode deviance as manageable risk.

Trusted Roles Lower Suspicion: Identification

Institutions do not merely act; they describe themselves. Institutional identities—executive, advisor, accountant, regulator, fiduciary—help pre-classify actors as trustworthy and make suspicion less likely. Operative form: Identification.

Stabilized self-descriptions organize expectation and shape how actions are interpreted. Friedrichs' analysis demonstrates that elite offenders often benefit from reputational capital: high-status identity elevates suspicion thresholds (Friedrichs, 2010). Black's (1976) theory of the behavior of law suggests that law responds differently to actors of varying status—elite identity can reduce the likelihood of aggressive prosecution, reinforcing stabilization. Internally, organizational culture may align employee loyalty with corporate narrative; critics may be marginalized as disloyal or incompetent. Identification thus functions as both shield and filter.

Boundaries Crossed Without Appearing Broken: Crossing

Legal and moral boundaries are crossed while preserving the appearance that no boundary has been crossed at all. Operative form: Crossing.

Crossing refers to movement across institutional domains—converting economic resources into political influence, transforming regulatory knowledge into competitive advantage, or leveraging public authority for private gain. White-collar crime frequently thrives at system intersections. Revolving-door employment between regulatory agencies and corporations blurs adversarial boundaries. Complex corporate structures may distribute activities across jurisdictions with varying enforcement intensity. Code-switching across domains allows actors to exploit structural gaps, because each domain applies its own criteria and what appears lawful in one context may facilitate evasion in another.

Suspicion Delayed, Diffused, or Neutralized: Void

The constitutive void of white-collar crime is the absence, delay, or neutralization of suspicion. Harm can be real and massive even while recognition remains fragmented or missing. Operative form: Void.

The void designates managed absence—what remains unmarked, undisclosed, or fragmented. Void production can occur through secrecy, non-disclosure agreements, technical opacity, or bureaucratic fragmentation. Information may exist but remain compartmentalized across agencies, such that no single actor possesses an integrated view. Void is not merely concealment; it is structural invisibility. When harms lack a clear focal victim or are distributed incrementally, mobilizing enforcement becomes difficult.

Configurations and Interaction

These operative forms rarely function independently. Distinction often requires supportive indication. Recursion intensifies boundary manipulation. Identification elevates suspicion thresholds, enabling void to persist. Crossing may reinforce re-entry by diffusing accountability across institutional domains. The grammar therefore invites configuration analysis: rather than asking whether a particular act constitutes white-collar crime, we ask how operative forms align to stabilize distortion. The following section applies this grammar empirically.

Empirical Illustrations: Three Institutional Settings, One Shared Mechanism

To demonstrate the explanatory reach of the meaning model, we examine three cases in which the shared mechanism—legitimacy organizing meanings that delay suspicion and conceal wrongdoing—is exhibited with particular clarity. The cases are selected not as a representative sample from which statistical generalization is intended, but as paradigm cases in the sense established by methodological discussions of case-based reasoning in organizational and historical sociology: instances chosen because they display the phenomenon under investigation in an unusually visible and well-documented form (Vaughan, 1992). Enron, Madoff, and Watergate differ in sector, scale, and historical moment, yet each exhibits the same underlying logic. This variation across context, combined with convergence in mechanism, supports the generalizability of the operative framework more persuasively than a set of similar cases would. Enron receives extended treatment as the fullest illustration of all seven operative forms working in concert. Madoff and Watergate are treated more briefly to establish that the mechanism is not sector-specific.

Enron: Fraud as Innovation

The Enron collapse offers a paradigmatic illustration of how operative forms converge to sustain distortion within a formally legitimate institutional environment. Organizational prestige, accounting complexity, and executive credibility helped normalize harm. The case has been widely analyzed as accounting fraud, executive corruption, and regulatory failure (Healy & Palepu, 2003; McLean & Elkind, 2003; United States Senate, 2002). What is less often specified is its communicative architecture.

At the level of distinction, special purpose entities (SPEs)—including Chewco, LJM1, LJM2, and the Raptor vehicles—were used to shift debt and underperforming assets off Enron's consolidated balance sheet while economic exposure remained substantively tied to the firm (United States Senate, 2002). Accounting rules permitted non-consolidation of SPEs if independent outside equity investment reached at least three percent of the entity's capital. Enron structured transactions to satisfy this technical threshold even when independence was compromised. The form of compliance was preserved; the substance diverged. Classification authority became a site of strategic action.

At the level of indication, Arthur Andersen issued clean audit opinions through fiscal year 2000. Credit rating agencies maintained investment-grade ratings through most of 2001. Prominent analysts continued favorable recommendations. These signals coordinated expectation among investors and counterparties. Gatekeepers occupied structurally ambivalent positions: Andersen's consulting relationships with Enron created economic interdependence; rating agencies relied on issuer-paid models; analysts operated within investment banks with underwriting interests. None of these dynamics required explicit collusion—structural incentives aligned indication with corporate narrative (Coffee, 2005).

Recursion operated through Enron's use of mark-to-market accounting for long-term energy contracts, under which projected future earnings were recognized immediately based on internal valuation models (Healy & Palepu, 2003). Each reported gain established a baseline for future performance. When underlying projects underperformed, new transactions were structured to offset losses and preserve reported earnings. The longer recursive cycles reproduced inflated performance, the more disruptive correction became.

Re-entry was visible when early warning signals were processed without structural reform. Sherron Watkins' August 2001 memorandum to CEO Kenneth Lay warned of potential implosion. External counsel reviewed transactions for compliance with applicable standards. The board examined but did not fundamentally restructure SPE relationships. Critique was absorbed procedurally, framed as manageable risk, and evaluated within existing classification frameworks—a textbook case of reflexive recoding.

Enron's identification as an innovative market architect transforming the energy sector elevated suspicion thresholds systemically. Complex transactions were interpreted as sophisticated innovation rather than warning signals. Executives were perceived as visionary rather than reckless. This identity influenced enforcement dynamics and internally discouraged dissent.

Crossing was evident in Enron's deep embeddedness within energy deregulation and financial innovation policy circles. Revolving-door dynamics between government and industry blurred adversarial boundaries. Domain entanglement complicated accountability and reinforced re-entry by distributing responsibility across institutions.

Finally, void was perhaps the most decisive operative form. Information relevant to Enron's exposure was dispersed across internal divisions, audit firms, rating agencies, regulators, and counterparties. SPE structures obscured consolidated liability. No vantage point synthesized the full configuration of risk. Collapse was triggered in October 2001 when Enron announced a $1.2 billion reduction in shareholder equity and disclosed off-balance-sheet liabilities. When indication fractured, the void collapsed: what had been compartmentalized became integrated through market reaction, and bankruptcy followed in December 2001.

The Enron case is not reducible to individual greed or accounting error. It was a configuration of operative forms: distinction redefined liability; indication stabilized credibility; recursion escalated commitment; re-entry absorbed critique; identification insulated perception; crossing diffused accountability; void fragmented visibility. The system did not lack documentation. It lacked structural transparency.

Bernie Madoff: Trust as Cover

The Madoff fraud illustrates the operative mechanism in a more condensed form. Reputation, exclusivity, and elite trust delayed suspicion and sustained the scheme for decades. Where Enron's architecture depended on organizational complexity across multiple operative forms, Madoff's relied above all on indication and identification.

Madoff's standing as a former NASDAQ chairman and prominent Wall Street figure constituted a near-impenetrable credibility signal. Investors—many of them sophisticated institutions—treated his consistent, implausibly smooth returns as marks of superior strategy rather than as warning signs. The fund's exclusivity deepened identification: investors valued membership in the scheme and were disinclined to scrutinize what appeared to be privileged access. Void operated through the deliberate simplicity of Madoff's paper trail—complex enough to obscure but simple enough to avoid drawing in sophisticated technical examination. The SEC received credible warnings on multiple occasions; re-entry absorbed these signals without structural consequence. Here, as in Enron, legitimacy organized meanings that systematically delayed recognition.

Watergate: Authority as Concealment

Watergate extends the framework into the domain of political misconduct, demonstrating that the mechanism is not limited to financial crime. Official power and institutional role were used to manage appearances and defer recognition. The Nixon administration's conduct illustrates how governmental authority can function as communicative cover.

Identification was central: the office of the presidency itself provided an epistemic shield. Crossing was systematic—executive authority was leveraged to obstruct legal process, direct intelligence resources, and influence enforcement. Re-entry was visible in the administration's repeated attempts to reframe incriminating revelations as politically motivated attacks or procedural disputes. Void was produced through deliberate information management, including the erasure of tape recordings and the compartmentalization of knowledge across White House staff. Recognition ultimately arrived not through institutional detection but through the collapse of internal loyalty and the persistence of journalistic and congressional investigation. The shared mechanism across all three cases is clear: legitimacy organizes meanings that delay suspicion and conceal wrongdoing.

Operative Dignity and Institutional Integrity

White-collar crime produces tangible harms: financial loss, unemployment, pension depletion, market destabilization. Yet the preceding framework suggests that elite deviance operates through a more fundamental layer of institutional life: the communicative conditions under which actors participate in organized systems. We use the term operative dignity to designate the integrity of those conditions.

Operative dignity is related to but distinct from several neighboring concepts. Trust, as treated in the white-collar crime literature, designates a relational condition between actors—confidence placed in another party that may be honored or betrayed. Procedural integrity, in the tradition of procedural justice, refers to the fairness of decision-making processes as experienced by participants. Legitimacy, in both Weberian and systems-theoretic usage, concerns the normative acceptance of authority. Operative dignity differs from each of these in that it is neither relational, experiential, nor normative in the first instance. It is structural: it designates the minimal communicative conditions under which the representational systems of institutions remain capable of orienting action. One can have legitimate authority, procedurally fair processes, and interpersonal trust, and still have operative dignity violated—as Enron illustrates, where formal compliance coexisted with systematic distortion of the representational infrastructure upon which investors, employees, and regulators depended.

This concept does not introduce a moral theory in the abstract. Rather, it names a structural requirement of institutional participation. Modern institutions depend upon representational infrastructures—financial disclosures, regulatory classifications, audit certifications, compliance reports, ratings, performance indicators—through which participants coordinate action. Investors allocate capital based on disclosed financial conditions. Employees make career and pension decisions based on corporate reporting. Regulators allocate oversight resources based on compliance documentation. In each case, participation presupposes that representational systems provide minimally reliable orientation.

When communicative architecture is systematically distorted, actors continue to act—but under degraded informational circumstances. Agency persists, yet it becomes structurally misaligned with underlying reality. Decisions made within distorted representational fields cannot be fully autonomous in the institutional sense, because the conditions of evaluation have been engineered. This reframing extends harm analysis beyond distributive loss to identify informational degradation as a structural injury.

The connection to the operative forms is direct. Identification shapes suspicion thresholds: when institutions stabilize high-status self-descriptions, interpretive filters adjust such that ambiguity is more likely to be read as sophistication rather than warning. Void amplifies this injury: fragmented oversight structures and compartmentalized information prevent integrated risk assessment. Enforcement asymmetry further compounds these dynamics; when boundary manipulation, signal alignment, and identity insulation intersect with uneven enforcement, distorted representation may persist longer than equivalent misconduct in lower-status contexts (Black, 1976).

From a policy perspective, operative dignity shifts the normative focus from punishing individual offenders to safeguarding institutional architecture. Traditional responses emphasize deterrence: increased penalties, enhanced prosecutorial capacity, whistleblower incentives. These measures remain important. But a communicative framework suggests complementary structural strategies—clarifying consolidation rules to reduce boundary engineering, strengthening the independence of audit and rating systems, reducing short-term performance incentives that intensify recursion, enhancing inter-agency information integration, and institutionalizing dissent channels that resist identity insulation.

We emphasize that operative dignity does not imply perfection. Institutional systems will always operate under complexity and uncertainty. The concept identifies a threshold condition: representation must not be systematically engineered in ways that prevent meaningful evaluation by participants. White-collar crime, in this light, represents a failure of structural integrity rather than simply a failure of character.

Conclusion: Order as Camouflage

White-collar crime is social harm produced through the manipulation of meaning under the cover of legitimate institutions. Legitimacy, trust, procedure, and institutional identity can become the very media through which wrongdoing is concealed and suspicion is delayed. White-collar crime is not merely elite deviance; it is a communicative process through which institutions can organize harm while preserving legitimacy. What appears lawful, technical, or routine may function as meaning cover. One of its deepest powers is not only that it causes harm, but that it manages suspicion. White-collar crime often uses order itself as camouflage.

A meaning model is more explanatory than trust-, occupation-, or status-based definitions because it specifies the communicative mechanisms by which harm is produced, concealed, and normalized. By centering institutional meaning-production, our framework moves beyond post hoc scandal analysis toward structural diagnosis. The grammar of seven operative forms—distinction, indication, recursion, re-entry, identification, crossing, and void—clarifies how representational systems can be configured in ways that delay recognition of harm. Operative dignity clarifies why such engineering matters beyond individual loss.

This reframing does not displace existing criminological theories. Strain illuminates performance pressures that animate recursive escalation. Opportunity theory clarifies how authority asymmetries facilitate boundary manipulation. Organizational research explains normalization processes that support re-entry. Regulatory scholarship identifies enforcement asymmetries that interact with identity and crossing. What the operative framework adds is structural precision: it identifies how these forces are mediated through classification systems, credibility signals, reporting cycles, reflexive review processes, domain entanglements, and informational fragmentation.

The empirical illustrations underscore the generalizability of the mechanism. Enron, Madoff, and Watergate are separated by sector, scale, and historical moment, yet each exhibits the same shared logic: legitimacy organizes meanings that delay suspicion and conceal wrongdoing. The cases differ not in mechanism but in which operative forms are most prominent and how they interact.

The broader implication is methodological as well as theoretical. Each operative form suggests measurable variables and comparative research pathways. Distinction can be examined through patterns of reclassification and restatement frequency. Indication can be analyzed through audit tenure and ratings persistence. Recursion can be modeled through performance volatility and escalation trajectories. Re-entry can be studied through regulatory settlements and internal investigation outcomes. Identification can be measured through reputational capital and enforcement differentials. Crossing can be mapped through revolving-door employment and political contribution networks. Void can be assessed through oversight fragmentation and disclosure lag metrics.

A criminology attentive to institutional meaning-production can therefore move beyond scandal analysis toward structural diagnosis—examining how classification design, signal independence, oversight integration, and identity narratives shape the stabilization of elite misconduct. White-collar crime endures not only because actors seek advantage, but because institutions provide communicative pathways through which distortion can become durable. Making those pathways visible is a necessary step toward reducing their structural vulnerability.

References

Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. Criminology, 30 (1), 47–87. https://doi.org/10.1111/j.1745-9125.1992.tb01093.x

Agnew, R. (2011). Toward a unified criminology: Integrating assumptions about crime, people, and society. New York University Press.

Benson, M. L., & Simpson, S. S. (2015). White-collar crime: An opportunity perspective (2nd ed.). Routledge.

Black, D. (1976). The behavior of law. Academic Press.

Braithwaite, J. (1985). To punish or persuade: Enforcement of coal mine safety. State University of New York Press.

Clinard, M. B., & Yeager, P. C. (1980). Corporate crime. Free Press.

Coffee, J. C., Jr. (2005). A theory of corporate scandals: Why the USA and Europe differ. Oxford Review of Economic Policy, 21 (2), 198–211. https://doi.org/10.1093/oxrep/gri012

Friedrichs, D. O. (2010). Trusted criminals: White collar crime in contemporary society (4th ed.). Wadsworth Cengage Learning.

Healy, P. M., & Palepu, K. G. (2003). The fall of Enron. Journal of Economic Perspectives, 17 (2), 3–26. https://doi.org/10.1257/089533003765888403

Luhmann, N. (1995). Social systems. Stanford University Press.

Luhmann, N. (2004). Law as a social system. Oxford University Press.

McLean, B., & Elkind, P. (2003). The smartest guys in the room: The amazing rise and scandalous fall of Enron. Portfolio.

Shapiro, S. P. (1990). Collaring the crime, not the criminal: Reconsidering the concept of white-collar crime. American Sociological Review, 55 (3), 346–365. https://doi.org/10.2307/2095761

Simpson, S. S. (2002). Corporate crime, law, and social control. Cambridge University Press.

Sutherland, E. H. (1940). White-collar criminality. American Sociological Review, 5 (1), 1–12. https://doi.org/10.2307/2083937

Sutherland, E. H. (1949). White collar crime. Dryden Press.

Tappan, P. W. (1947). Who is the criminal? American Sociological Review, 12 (1), 96–102. https://doi.org/10.2307/2086496

United States Senate Permanent Subcommittee on Investigations. (2002). The role of the board of directors in Enron's collapse. U.S. Government Printing Office.

Vaughan, D. (1992). Theory elaboration: The heuristics of case analysis. In C. C. Ragin & H. S. Becker (Eds.), *What is a case? Exploring the foundations of social inquiry* (pp. 173–202). Cambridge University Press.

Vaughan, D. (1996). *The Challenger launch decision*. University of Chicago Press.

Stockholm Syndrome: Origins, Evidence, and Institutional Legitimation

Keith Robert Head

Capella University

Abstract

This article examines whether Stockholm syndrome, a term that emerged from the 1973 Norrmalmstorg bank robbery, is supported by the available evidence. Using primary sources and a review of the literature, it finds that the psychiatrist who pioneered the label never assessed the hostages, that the hostages' behavior may reflect reasonable responses to danger, and that the term was introduced after a hostage publicly criticized how police handled the situation. Drawing a distinction between the cultural reach of Stockholm syndrome and the clinical record behind it, the article explores the consequences of this gap for psychology and its neighboring fields. The term continues to appear in textbooks and applied settings despite this evidentiary gap. Scholarship also suggests that the label has often been applied along gendered lines, including to women whose responses were characterized as irrational. The term's wide circulation may have narrowed what questions get asked about captivity bonding and kept explanations like trauma bonding from receiving consideration. From this, the article develops a broader account it terms circular institutional legitimation, in which a construct gains apparent scientific authority through successive institutional adoptions that substitute for empirical validation.

Keywords: Stockholm syndrome, Norrmalmstorg robbery, psychiatric diagnosis, victim pathologization, trauma bonding, construct validity, circular institutional legitimation

Introduction

Most psychological constructs enter professional use through a familiar sequence. A pattern of behavior is observed, a hypothesis is proposed, studies are designed to test it, and if the evidence holds, the concept earns its place in clinical practice. Stockholm syndrome did not follow this sequence. It was named on a news broadcast by a psychiatrist who had not examined the hostages, taken up by law enforcement agencies seeking a framework for hostage negotiation, introduced into American courtrooms during a high-profile kidnapping trial, and later formalized by a clinician working in part from secondhand accounts. No controlled empirical study appears at any point in this chain. Yet it persists in clinical training, forensic testimony, and popular understanding as if it had passed through that process.

The question of how this happened is not unique to Stockholm syndrome, but the case is unusually well suited to answering it. The concept's origins are traceable to a single event, the 1973 Norrmalmstorg bank robbery in Stockholm, and its subsequent path through institutions can be followed with a specificity that broader diagnostic categories do not permit. This paper uses that traceability to examine a problem that extends well beyond one contested label. How does a concept born outside research enter clinical and legal use, and once embedded, why does it resist correction even when its evidentiary basis is contested?

Stockholm syndrome, otherwise known as Norrmalmstorg syndrome (Norrmalmstorgssyndromet in Swedish), while not formally defined, is generally considered the psychological tendency of a hostage to bond with, identify with, or sympathize with their captors. From its inception the definition has been unstable. Bejerot (1974) described it as hostages' paradoxical emotional attachment to their captors. The FBI later defined it as a psychological response in which captives develop positive feelings, sympathy, or emotional bonds toward their captors, sometimes to the point of defending them or identifying with their cause (De Fabrique et al., 2007). Although the syndrome has never been included in the Diagnostic and Statistical Manual of Mental Disorders or the International Classification of Diseases, mental health professionals and law enforcement personnel have applied this label extensively when attempting to explain trauma victims' responses.

That application continues in settings where it carries real consequences. Forensic psychiatrists in sex trafficking prosecutions are still called on to explain survivor behavior through Stockholm syndrome despite the absence of any diagnostic framework behind the term (Zhang & Datta, 2022), and as recently as 2018 a New York court heard expert testimony that invoked the Norrmalmstorg case by name to assess a trafficking survivor's credibility (Justia, 2018). The concept has also been written into federal policy, with the U.S. State Department treating Stockholm syndrome and trauma bonding as interchangeable in its official guidance on human trafficking (U.S. Department of State, 2020). Meanwhile, clinicians continue to treat it as though it were an established diagnostic entity, with a 2025 case report describing monitoring for the emergence of Stockholm syndrome during PTSD recovery in a domestic violence survivor as a routine clinical consideration (Alshwaiheen et al., 2025). That application has been criticized as gendered. The label appears to have been applied disproportionately to women, and in the foundational case, the term was introduced after a female hostage publicly criticized the police response, a sequence later commentators read as recasting her objections as pathology. As the concept migrated into the literature on intimate partner violence during the 1990s, it carried an explanatory logic that could position women's responses to abuse as symptoms of a syndrome rather than as intelligible reactions to dangerous circumstances.

This paper has two aims. It examines whether the available evidence supports Stockholm syndrome as a clinical construct, and it follows how the term came to be applied as a syndrome and how it spread into clinical and legal use. Drawing on primary sources, published literature, and the accounts of those directly involved, it asks what the term's survival tells us about the way psychiatric knowledge is produced, circulated, and maintained. The analysis proposes that Stockholm syndrome's trajectory represents a recognizable pattern in which constructs acquire apparent scientific standing not through empirical validation but through successive institutional adoptions that collectively substitute for it, a process this paper terms circular institutional legitimation. The implications reach beyond this single case to any construct whose authority rests on the same kind of unexamined circulation. The evidentiary critique that follows is therefore offered not as an end in itself but as the means of showing how Stockholm syndrome acquired its standing.

Methods

A narrative review supported by selected primary sources was used to assess the 1973 Norrmalmstorg bank robbery and the subsequent development of Stockholm syndrome. Rather than provide an extensive systematic review of all literature on the topic, this paper aims to determine how the concept evolved from emergence into both clinical and public domains, and whether existing literature provides support for the continuing use of Stockholm syndrome as a psychological model. Sources were gathered through searches of PubMed, PsycINFO, PsycARTICLES, SocINDEX with Full Text, Criminal Justice Database, PILOTS, ScienceDirect, SAGE Journals Online, and Scopus for publications from 1973 to 2025, using search terms, alone and in combination, including "Stockholm syndrome", "Norrmalmstorgssyndromet", "Norrmalmstorg", "Bejerot", "hostage psychology", "hostage negotiation", "trauma bonding", and "appeasement". Because critique of the concept was as relevant as its endorsement, terms tied to competing explanations were searched alongside the concept's own terminology so that revisionist and dissenting work would also be returned. The reference lists of retrieved articles were then screened, and forward citation tracing was used to capture earlier and later work that the database search did not return. Primary materials on the foundational case were identified separately, since many predate database indexing or fall outside it; these included contemporary news coverage, published interviews, memoirs, the relevant FBI Law Enforcement Bulletin articles, and the Patricia Hearst trial record.

The literature reviewed in this paper was selected based upon relevance to five guiding research questions: (1) How did the concept emerge? (2) How was the concept defined and disseminated? (3) Does the literature demonstrate clinical validity? (4) Are there alternative frameworks that can be used to describe the behaviors typically referred to as Stockholm syndrome? (5) By what process did the concept acquire and retain institutional authority despite that evidentiary record? A source was retained when it bore directly on the origin, clinical application, or critique of the concept as it related to these questions, and was excluded when it referred to the term only in passing. No restriction was placed on the form of the work, so empirical studies, reviews, conceptual papers, interviews, books, and historical and sociological analyses were all eligible. Peer-reviewed literature as well as selected historical documents were utilized in this review. Peer-reviewed scholarship was treated as the basis for assessing clinical validity, while these historical materials were used to establish what was claimed about the concept and when.

Conclusions were reached through narrative synthesis. The reviewed material was arranged chronologically to trace how the concept moved from its coinage into law enforcement training, the courtroom, formal clinical definition, and later academic use, and each stage was examined for whether its adoption rested on empirical validation or on prior institutional use. Where these sources gave competing accounts of what occurred, the firsthand testimony of those directly involved was given precedence over third-party and official accounts on questions concerning their own conduct and motives. Claims that recurred across independent sources were treated as better supported than those traceable to a single interested party, and points on which the documented history of the event remains contested were noted as such rather than reconciled.

Historical Background

On August 23, 1973, 32-year-old Jan-Erik Olsson, a convicted safecracker who had failed to return from a prison furlough, entered the main branch of Sveriges Kreditbank at Norrmalmstorg square in central Stockholm armed with a loaded submachine gun. He fired into the ceiling, triggering the bank's silent alarm. When Detective Inspector Ingemar Warpefeldt responded, Olsson shot him in the right hand. Olsson took three female employees hostage, binding them with rope: 23-year-old Kristin Enmark, 21-year-old Elisabeth Oldgren, and 31-year-old Birgitta Lundblad. His demands included three million Swedish kronor, weapons, and the release of his former prison cellmate Clark Olofsson, then serving a sentence for armed robbery and acting as an accessory in the murder of a police officer. The government refused to let Olsson leave the bank with the hostages but delivered the ransom money and Olofsson. For two days, police misidentified Olsson as Kaj Hansson, another escaped prisoner, and sent Hansson's 16-year-old brother to the bank to plead with the man they believed to be Hansson. Olsson, who did not recognize the boy, assumed it was a ruse and fired at him. The teenager fled to safety. Olsson's true identity was only established when a detective recognized his voice during a radio broadcast. Olofsson arrived at the bank shortly after four on Thursday afternoon. He discovered a fourth employee, 25-year-old Sven Säfström, hiding in a stockroom. The group moved into the bank's ground-floor safe-deposit vault. On Saturday morning, a police officer crept down the staircase and locked the outer vault door from the outside. Olsson barricaded the inner door and planted an explosive charge in a ventilation hole (King, 2020).

Police sought the advice of Nils Bejerot, a prominent psychiatrist who advised throughout the crisis that the longer Olsson spent with the hostages, the more likely he was to capitulate. This guided the police strategy of attrition. Police laced beer with a soporific before lowering it to the vault, but the hostages refused it because the beer looked flat and the caps were askew (King, 2020). On Thursday evening, Olsson phoned Prime Minister Olof Palme and threatened to kill the hostages, seizing Oldgren by the throat and beginning a countdown. He stopped at fifteen seconds and hung up. The following day, Enmark called Palme and spoke for forty-two minutes, pleading to be allowed to leave with the robbers. She said she trusted Olofsson and Olsson completely and that the police, not her captors, were what frightened her. Palme told Enmark that she would not be leaving the bank and that she should content herself with dying at her post. Enmark was appalled, responding that she had no intention of being a dead hero. Palme refused to negotiate further (Enmark & Wesslén, 2020).

Throughout the standoff, the captors showed intermittent care toward the hostages. Olsson consoled Lundblad when she wept over her children. He gave Enmark a bullet as a keepsake. He wiped tears from Oldgren's face and draped his jacket around her shoulders when the cold woke her. When Oldgren suffered claustrophobia, Olsson attached her to a rope and let her walk outside the vault. When Enmark had a nightmare and awoke screaming, Olofsson soothed her. At the same time, Olsson threatened to shoot Säfström in the leg to pressure police. On Monday, August 27, to deter a gas attack, Olsson and Olofsson placed ropes around the necks of all four hostages and had them stand before safe-deposit cabinets with the ropes knotted to drawer handles, so that if gas rendered them unconscious they would be strangled as they fell. The hostages stood in these nooses for hours (Olofsson & Grimwalker, 2023).

On the evening of Tuesday, August 28, despite threats to kill the hostages if they did so, police pumped tear gas through holes drilled in the vault ceiling. Olsson surrendered. When officers called for the hostages to come out first, all four refused. Enmark shouted, "No, Jan and Clark go first, you'll gun them down if we do!" The hostages and captors embraced, the women kissing their captors, Säfström shaking hands. As Enmark was placed on a stretcher, she saw the handcuffed Olofsson being roughed up by police and screamed that he hadn't done anything, trying to defend him (King, 2020; Lang, 1974).

Aftermath of the Case

Following the Norrmalmstorg robbery, the hostages' conduct throughout the ordeal and in the period that followed was questioned by both Swedish officials and the general public, including their expression of empathy for their captors and criticism of the police response. During the final police assault, they refused to leave the vault before their captors because they feared the police would execute the men if given the opportunity. This protective stance was captured on live television as the victims embraced their captors and Kristin Enmark called out to Olofsson that she would see him again. The case drew widespread public attention as the individuals who had been held at gunpoint for six days appeared to align themselves with their victimizers rather than their rescuers (Adorjan et al., 2012).

In the immediate aftermath of her release, Kristin Enmark went on Swedish radio and publicly criticized the police, along with Nils Bejerot, the police psychiatrist who had served as a consultant throughout the standoff. She accused him of endangering the hostages' lives by recommending aggressive tactics like the use of tear gas and drilling through the vault ceiling (Enmark & Wesslén, 2020). Bejerot then characterized the hostages' reactions as a product of brainwashing and introduced the term Norrmalmstorgssyndromet, which Åse (2015) reads as recasting the women's rational responses as pathology (Bejerot, 1974). The timing and context of this coinage are important. Bejerot introduced the theory publicly only after her criticism, a sequence several commentators read as a direct response to it. Olofsson himself later dismissed the concept, writing that the syndrome "was never a syndrome" and that he was "just nice and we liked each other," attributing it to academics like Bejerot who "wrote dissertations" about what he considered ordinary human interaction (Olofsson & Grimwalker, 2023).

During the subsequent legal proceedings, tension between the hostages and the authorities continued. Although it is often misreported that they refused to testify, they did take the stand reluctantly but acted as difficult witnesses for the prosecution. Their testimony was often characterized by equivocation to avoid incriminating the captors. Kristin Enmark, for instance, admitted in later accounts that she intentionally lied during the trial when asked if she had seen Olofsson holding a weapon, as she felt he had saved her life and did not want him to be punished (Enmark & Wesslén, 2020). This limited cooperation sat awkwardly against the prosecution's case and left Bejerot's label as the readily available public explanation for the victims' actions. Through this coverage, Norrmalmstorgssyndromet circulated widely in the Swedish media and among the public.

Evolution of the Concept

However, the theory of Norrmalmstorgssyndromet did not end with Enmark, Bejerot, or Sweden. On February 4, 1974, less than one month after the Norrmalmstorg trial began, Patricia Hearst, a 19-year-old Berkeley art history student and heiress to the Hearst publishing fortune, was kidnapped from her apartment by members of the Symbionese Liberation Army (Adorjan et al., 2012). This event would change the trajectory of the term. Two months later, surveillance footage captured her participating in a bank robbery alongside her former captors. Psychiatrist Frederick J.

Hacker, who had advised the Hearst family, began using the term Stockholm Effect to describe her apparent transformation (Hacker, 1977). Following her arrest in September 1975, Hearst came to be considered a textbook example of Stockholm syndrome (Cerny et al., 2010). The Hearst case established Stockholm syndrome in both legal and popular consciousness. At her 1976 trial, the defense invoked Stockholm syndrome to argue that Hearst had been coerced and psychologically manipulated into criminal activity. The jury rejected this defense, and Hearst was convicted of bank robbery (Rahaim & Matusitz, 2020). Nevertheless, the trial brought unprecedented public attention to the concept and established a precedent for Stockholm syndrome as a potential legal defense, a strategy that would be attempted in subsequent cases involving allegations of coerced criminal participation.

The Hearst trial illustrates how legal proceedings conferred credibility on the concept without requiring empirical validation. F. Lee Bailey's defense team invoked Stockholm syndrome as the framework for Hearst's transformation, yet the expert witnesses on both sides testified about coercion, brainwashing, indoctrination, and psychological manipulation in general terms rather than presenting data specific to Stockholm syndrome as a validated construct. The court heard over 200 hours of psychiatric testimony, but none of the experts were asked to produce, nor did they offer, empirical evidence that Stockholm syndrome had been established as a replicable phenomenon with agreed-upon diagnostic criteria. The prosecution's expert Joel Fort was not a psychiatrist, and his standing as an expert on coercive persuasion was contested (Alexander, 1979). The defense's Robert Jay Lifton acknowledged reaching conclusions before reviewing the evidence in the case (Ewing & McCann, 2006). The trial treated Stockholm syndrome as though it were an accepted clinical entity because it had already been named, discussed in FBI training contexts, and referenced in media coverage. What the trial produced was not scientific validation but legal visibility: a high-profile courtroom in which the concept was debated as though its existence were settled and only its applicability to Hearst remained in question.

By the mid-to-late 1970s, the law enforcement community had begun to embrace Stockholm syndrome as a practical concept for crisis negotiation. Harvey Schlossberg, a New York police psychologist, claimed to have coined the term to help officers understand hostage dynamics (Schlossberg & Freeman, 1974). He and Captain Frank A. Bolz Jr. developed the world's first hostage negotiation program, successfully resolving 285 incidents without loss of life by 1982 (Grubb, 2023). In his influential 1979 paper in the FBI Law Enforcement Bulletin, Thomas Strentz characterized Stockholm syndrome as an automatic, unconscious emotional response to victimization involving regression to a childlike dependency state. Strentz explained that the perpetrator's supreme power transforms each breath into a perceived gift. He warned officers to expect hostility from hostages themselves, describing this as "the price that law enforcement must pay for a living hostage" (Strentz, 1979). This framework became foundational to FBI hostage negotiation training at Quantico.

Strentz's paper, the Bureau's central reference for hostage situations, cited the Norrmalmstorg case and Ochberg's emerging theoretical framework as its basis. The FBI Law Enforcement Bulletin was an internal agency publication, not a peer-reviewed journal, and no record indicates that the concept underwent independent empirical review before its integration into the national training curriculum. The Bureau adopted the concept on the strength of Bejerot's label and Ochberg's theorizing rather than on the basis of controlled research, yet the FBI's institutional prestige meant that its adoption was subsequently treated by other agencies and by academic sources as evidence that the concept had a credible foundation. By 1982, the NYPD model developed by Schlossberg and Bolz had been brought to Quantico and expanded into a national program, with Stockholm syndrome embedded in the curriculum. The FBI's own HOBAS database would not produce data on the syndrome's prevalence until 1999, meaning the concept was taught as established for two decades before any internal data on its prevalence existed (Fuselier, 1999).

This FBI training course was built upon the work of Frank M. Ochberg, a Johns Hopkins and Stanford-trained psychiatrist at the National Institute of Mental Health. During this same period, he studied hostage situations in Europe as part of his role on the National Task Force on Disorders and Terrorism. Ochberg was drawn to Norrmalmstorg because it was a highly publicized hostage crisis. Through this work he came into contact with FBI Special Agent Conrad V. Hassel, with whom he later collaborated on the study of hostage victimization. In 1982, Ochberg articulated what became the most widely accepted definition of Stockholm syndrome, identifying three basic components: the hostage develops some form of attachment to the captor; the attacker reciprocates this sentiment to the point of wanting the hostage to survive; and both captor and hostages come to distrust police and authorities (Ochberg & Soskis, 1982). From the New York Police Department and the FBI Training Division, the concept spread to police departments across the United States and eventually to more than two dozen countries worldwide.

Ochberg's account of how he arrived at his definition is instructive. In interviews he has described writing a memo to the FBI defining Stockholm syndrome based on his work with the National Task Force on Disorders and Terrorism and his subsequent consultations in Europe (Aponte, 2020). He has stated that he debriefed many people who had been held hostage during the wave of terrorism-related incidents in the 1970s, including hostages held by Moluccan separatists in the Netherlands (King, 2020). However, Ochberg's definition of Stockholm syndrome drew specifically on the Norrmalmstorg case, an event already interpreted through Bejerot's framework.

This definition became canonical in law enforcement and eventually in academic literature. When Ochberg wrote his 1982 definition identifying three core components, he cited no controlled data, no validated instruments, and no peer-reviewed research establishing that these components co-occurred with any consistency. What he offered was a clinically plausible model derived from anecdotal observation, which the FBI and Scotland Yard then adopted for its tactical utility in the absence of empirical testing.

During the 1990s, researchers began applying the term beyond hostage situations to encompass abusive relationships, particularly those involving intimate partner violence. Graham et al. (1995) identified four precursors: perceived threat to survival, perception of small kindnesses from the captor, isolation from other perspectives, and perceived inability to escape. This extension built upon D. G. Dutton and Painter's (1981) concept of traumatic bonding in relationships characterized by intermittent abuse. This conceptualization suggested that the psychological mechanisms underlying Stockholm syndrome might operate in a variety of contexts beyond traditional hostage situations, including child abuse, cult involvement, and controlling romantic relationships (Jülich, 2005; West & Martin, 1996). Since then, the term has continued to be applied across these contexts.

Origins of the Concept

The concept's origins have been controversial from the outset. This contention stems largely from the dual role assumed by the psychiatrist who defined the condition, raising questions about professional boundaries. Bejerot served as the psychiatric consultant advising the Stockholm police throughout the 1973 Norrmalmstorg siege, placing him at the center of the police maneuvers that the hostages identified as a primary source of their fear and trauma (Lang, 1974). When hostage Kristin Enmark publicly criticized both the police operation and Bejerot by name on live radio, he subsequently labeled her reactions as symptomatic of a new condition he termed Norrmalmstorg syndrome. This response raises a potential conflict of interest, since Bejerot claimed diagnostic authority over responses to an operation he had helped direct. His institutional alignment with law enforcement meant that any assessment of hostage behavior would implicate his own tactical recommendations, raising questions about independent judgment.

Bejerot's conclusions about the hostages also rest on a significant methodological limitation, namely that he did not conduct clinical interviews with the individuals he described. By formulating a construct without direct evaluation or consent, Bejerot departed from core standards of psychiatric practice and risked reducing complex human responses to pathological categories based on external interpretation. The lack of empirical validity is further compounded by the fact that Bejerot advised police to seal the hostages in the vault, a strategy the hostages identified as their primary source of fear, which raises the question of whether the conditions he later described were in part produced by the strategy he advised (Bejerot, 1974; Enmark & Wesslén, 2020). This pattern of diagnosing without evaluation extended to Ochberg, who later formalized the syndrome based on secondhand accounts.

The conduct of these professionals departs from longstanding principles of medical practice regarding examination and the avoidance of harm. The American Psychiatric Association adopted the Goldwater Rule in 1973, the same year as the Norrmalmstorg crisis, stating that it is unethical for a psychiatrist to offer a professional opinion about an individual without having conducted a personal examination and obtained proper authorization (Appelbaum, 2017; Kroll & Pouncey, 2016). The American Psychological Association's Ethics Code, first published in 1953 (Stricker, 2010), has similarly required that assessments be grounded in adequate data and direct examination, and that professionals avoid causing harm through the misuse of their professional influence (Binder & Thompson, 1995). While these codes were not binding on a Swedish psychiatrist, they reflect professional standards that were widely recognized at the time. Bejerot publicly characterized individuals who had declined to be evaluated, while serving as a participant in the operation being criticized. Ochberg built on that foundation without correcting its methodological deficiencies. Offering diagnostic claims without examination or consent and while holding a potential conflict of interest runs counter to the principles of beneficence, nonmaleficence, integrity, and justice that have long governed both professions. These departures from basic psychiatric procedure leave Stockholm syndrome with a contested origin, since key conclusions behind it were reached without the hostages' consent and by a professional whose own conduct was part of what they were objecting to.

Clinical Validity

One of the most significant concerns with Stockholm syndrome is whether the concept satisfies the requirements of a syndrome at all. A syndrome, by definition, requires a recognizable pattern of co-occurring symptoms that presents with some consistency across individuals exposed to comparable conditions. The behaviors attributed to Stockholm syndrome do not meet this basic threshold. The FBI's Hostage Barricade Database System, drawing on over 1,200 federal, state, and local incidents, found that fewer than 8% of hostage victims displayed any behaviors associated with the alleged syndrome, a figure that dropped to roughly 5% once cases of simple frustration with law enforcement were excluded (Fuselier, 1999). That more than a third of identified cases could be attributed to ordinary irritation with authorities rather than any psychological bond with a captor suggests that the construct's boundaries are poorly defined and that even its limited empirical footprint may rest on inflated and inconsistent categorization.

No standardized or validated diagnostic criteria for Stockholm syndrome have ever been established. The concept itself has been defined in different ways by different authorities, ranging from Bejerot's initial characterization of the hostages as brainwashed to his later formulation of a paradoxical emotional attachment, to the FBI's description of captives developing positive feelings and sympathy for their captors, to Ochberg's three-component model involving attachment, reciprocation, and mutual distrust of authorities. Various researchers and commentators have proposed different sets of features or precursors, from Graham et al.'s (1994) four-factor model to looser formulations based on the presence of positive feelings toward a captor, but none of these have been subjected to the kind of rigorous testing that would produce agreed-upon thresholds for identification. This leaves the construct without any empirically sound method of verification. The term itself has also been applied inconsistently across markedly different contexts, from brief hostage incidents to prolonged domestic abuse to cult membership, with no clear rationale for why such varied situations should produce the same underlying condition (Namnyak et al., 2008). Without common criteria, what constitutes a "case" of Stockholm syndrome depends heavily on who is doing the labeling and what behaviors they choose to include, which is not how psychiatric classification is meant to work.

The empirical literature offers no meaningful support for Stockholm syndrome as a distinct psychological phenomenon. The concept originated in media coverage and was subsequently adopted into law enforcement training and popular culture, but this trajectory from journalism to jargon did not follow the usual process by which clinical constructs are tested, replicated, and refined through peer-reviewed research (Bailey et al., 2023). Systematic reviews of Stockholm syndrome have done more to confirm the absence of a research base than to establish Stockholm syndrome as a valid construct. Namnyak et al. (2008) found only twelve qualifying papers, most of which were case reports rather than empirical studies with controlled designs. The comparative analysis itself drew on just five high-profile cases sourced from media accounts, three of which involved minors whose responses to captivity cannot be meaningfully compared to adult behavior given fundamental differences in attachment, cognitive development, and emotional regulation. The four overlapping features identified across these cases are difficult to interpret given the small, non-representative sample and the strong likelihood of reporting and publication bias. Rizo-Martínez (2018), analyzing the literature, found no valid diagnostic criteria for the syndrome and attributed the patterns it identified to information bias. When systematic reviews must rely on case reports and theoretical proposals, compare child and adult responses, and still surface no valid diagnostic criteria because the clinical literature barely exists, it raises the question of whether there is enough evidence to validate the concept at all.

Efforts to measure the construct with a dedicated scale show the same problem. Graham et al. (1995) developed the most widely used such instrument, and the psychometric record complicates the claim that it identifies a distinct condition, since a scale identifies a real condition only if its underlying structure stays stable when applied to new groups, and Graham's scale has not held such a structure across later samples. When other researchers applied the same items to new samples, the structure changed, with the three components Graham identified becoming two in Obeid and Hallit (2018), drawn from 30 women, and seven in Rahme et al. (2020), drawn from more than 1,600; Rahme et al. reached seven only by discarding five items that did not fit, with no earlier validated version to check against, suggesting the structure was fitted to the sample rather than confirmed against a known one. Although internal consistency was high in both ($\alpha = .94$ and $\alpha = .97$), this shows only that the items cohere, not that they mark a distinct condition, and the scale correlated weakly with abuse and more strongly with general stress, indicating that it captured general distress rather than anything specific to a captor bond.

The same instability holds across samples drawn from different populations. A French-language validation administered to women in Cameroon retained only sixteen of the original forty-nine items, discarding thirty-three for inadequate loadings (Mouafo & Lontio, 2024), and a Spanish-language validation in Mexican women recovered the three factors only after proposing a reduced version of the scale (Santoyo Telles & Rizo Martínez, 2019). A mediation study in a Pakistani sample reached the same conclusion from the opposite direction, finding that the Core Stockholm Syndrome component did not mediate the relationship between intimate partner violence and psychological distress while a general psychological-damage component did (Ahmad et al., 2018). Across these samples the explanatory work was done by dependency, schema activation, and traumatic bonding, mechanisms that require no syndrome, with the authors retaining the label while their own findings accounted for the behavior without it.

It is therefore unsurprising that Stockholm syndrome has never been included in the Diagnostic and Statistical Manual of Mental Disorders (DSM), the primary classification system used in American psychiatric practice, nor in the International Classification of Diseases (ICD) used internationally. Nothing in the published record indicates that its absence reflects oversight or pending review. The concept has not met the evidentiary threshold that either system requires for inclusion (Namnyak et al., 2008), and no formal proposal for recognition appears in the literature. After more than five decades, the gap between Stockholm syndrome's cultural prominence and its absence from clinical classification remains unresolved.

Adaptive Responses to Threat

The concept of Stockholm syndrome may pathologize behaviors that can also be understood as adaptive survival responses. When a captive cooperates with, attends to, or appears emotionally connected to their captor, this does not necessarily indicate a disordered psychological bond. Bailey et al. (2023) proposed replacing the term with "appeasement," grounded in Polyvagal Theory, which explains these behaviors as reflexive neurobiological processes in which the autonomic nervous system activates the social engagement system under conditions of extreme threat. The function of this activation is to co-regulate with the captor, calming them and reducing the risk of injury or death. Labeling this process as a syndrome may reframe a survival response common to mammals as a form of psychological dysfunction.

The syndrome label may conflate the appearance of emotional connection with genuine affection, without fully accounting for what the behavior accomplishes in context. Enmark's own words illustrate this: "many hours after we were let out of the vault, I met a psychiatrist named Bergman whose first question was if I was in love with Clark Olofsson" (Enmark & Wesslén, 2020, p. 75). A captive who mirrors a captor's emotional state, expresses concern for their well-being, or avoids antagonizing them is not forming a bond in any meaningful relational sense. These behaviors serve a concrete, functional purpose in managing the threat level of an unpredictable person who holds total control over the captive's survival. The observable output may resemble affection or loyalty, but the underlying process is one of threat management. On this account, the syndrome label takes the surface presentation for the substance, interpreting behavior that serves to manage threat as evidence of sincere psychological attachment.

The syndrome label also imposes a pathological framework onto what may not require clinical explanation at all. Given the previously noted rate of only 5% to 8% of hostages displaying these behaviors, and given the absence of validated diagnostic criteria, these responses are arguably better understood as natural variation in human temperament and stress response. Some individuals are dispositionally higher in empathy, agreeableness, or affiliative behavior under threat (Taylor, 2012). Rather than requiring a pathological explanation, the responses of this small minority of captives may simply reflect the expected distribution of personality traits and coping styles across any population placed under duress.

It is also worth considering that in the concept's foundational case, the hostages' responses had a clear rational basis. Police made several missteps, including misidentifying the captor and sending in the teenage brother of the man they had wrongly assumed was inside, provoking Olsson to fire rounds and become significantly more agitated. Authorities later attempted to pump tear gas into the vault where the hostages were confined, causing panic among the captives who feared suffocation. When hostage Kristin Enmark was granted a phone call with Prime Minister Palme and pleaded for her life, he told her she should content herself with dying at her post, a remark Enmark took to mean that the hostages' survival came second. In this context, the hostages' stated preference for the captors over the police can be read not as paradoxical but as a straightforward assessment of where the greater threat lay. What was subsequently labeled a syndrome may have simply been an accurate, real-time reading of a situation in which institutional decisions repeatedly increased the danger to the hostages. Olofsson's own account of the events inside the bank makes a similar claim, stating that the standoff would never have produced the conditions later attributed to the syndrome had Palme not refused to allow a negotiated exit (Olofsson & Grimwalker, 2023). Enmark herself rejected the label entirely, stating that she was not sick and had not suffered from any syndrome, and that her actions in the vault were driven solely by the need to survive (Enmark & Wesslén, 2020).

Gender and the Label

Stockholm syndrome, as a diagnostic concept, can be read as a product of its cultural moment, disproportionately applied to women in ways that echoed gendered scripts about irrationality and emotionality. As Hill (2020) puts it, Stockholm syndrome is "a dubious pathology with no diagnostic criteria" that is "riddled with misogyny and founded on a lie" (Hill, 2020, p. 57). Such critiques position Stockholm syndrome as a gendered label, on the grounds that it has often been applied to women and draws on cultural scripts about feminine irrationality. Jameson (2010) traces how Stockholm syndrome relies on historically rooted assumptions about women's susceptibility to influence, connecting the concept to nineteenth-century discourses of hysteria and suggestibility. Under this critique, the syndrome positions emotional responsiveness not as strategic but as evidence of cognitive failure, reproducing a gendered logic in which women's relational capacities are recast as vulnerabilities that undermine their credibility.

Åse's (2015) analysis of the original Stockholm case suggests the label served masculinist state protection. The Swedish state's authority as protector had been undermined by a crisis it mishandled, and the women hostages' public criticism of police tactics threatened to make that failure visible. Åse argues that by characterizing the women's objections as symptoms of a syndrome, institutional actors restored a gendered order in which the state protects, and women receive protection. Enmark's own account is consistent with this reading. Reflecting on the aftermath, she writes: "They tried to take advantage of us ... to hide their own prejudices, mistakes, and the fact that they so miserably failed to protect us. The fact that they also treated us as mentally ill makes it even worse ... they thought that it's easy to blame us young girls" (Enmark & Wesslén, 2020, p. 62). What Åse theorizes as the regendering of women into dependent victims, Enmark describes as an infantilizing response, with young women treated as less capable of credible testimony.

Graham et al.'s (1994) theory of Societal Stockholm syndrome reinforces this structural critique from a different angle. They argue that the bonding behaviors attributed to individual hostage victims are structurally identical to the survival strategies women adopt under patriarchy at large. These behaviors are described as responses to the four precursor conditions of Stockholm syndrome, which are perceived threat to survival, perceived inability to escape, isolation from nonmale perspectives, and perceived kindness from men. The theory describes femininity as "a set of behaviors that please men (dominants) because they communicate a woman's acceptance of her subordinate status" and explicitly states that "feminine behaviors are survival strategies" (p. 237). It suggests that women's bonding to men, women's femininity, and women's heterosexuality may be understood as adaptive responses to male violence and captivity. The theory further notes that psychiatry tends to "pathologize victims' survival responses to unhealthy interpersonal environments" (p. 60), treating these as individual disorders rather than as responses to structural conditions of male domination.

Wade (2015) ties this together through research built on interviews he conducted with Kristin Enmark. He found her actions during the event were prudent, courageous resistance, not symptoms of psychological bonding. He argues that the invention of Stockholm syndrome entailed an abuse of male-state power that transformed Enmark's resistance into pathology. Enmark was responding to both the hostage takers and a disorganized and dangerous police-state response while protecting and maintaining solidarity with her fellow captives. On Wade's account, when she criticized the police and state response, Bejerot responded with Stockholm syndrome to silence her. He argues that Swedish media reinforced Bejerot's diagnosis because Enmark did not conform to expected scripts of victimhood. Her coherent and self-possessed demeanor was treated as suspicious rather than as a trauma response. A senior Swedish police official told Wade (2015) that the concept was fundamentally gendered, noting that it was easy to silence a woman simply by saying these things. In his account, the invention of Stockholm syndrome amounted to an abuse of male-state power, and the term itself is one of many closely related concepts that blame victims, mostly women, while denying the reality of their resistance and ignoring the central role of social responses. This aligns with the four discursive operations identified in accounts of violence, which are concealing violence, mitigating perpetrator responsibility, concealing victim resistance, and blaming or pathologizing victims (Coates & Wade, 2004). Read this way, Stockholm syndrome can be understood as performing all four operations simultaneously, functioning as a mechanism for discrediting women who refuse to remain silent about violence and institutional failures.

Competing Perspectives and Continued Uses

Despite these concerns, Stockholm syndrome has remained useful across clinical and forensic settings because it offers a compact way to describe victim behavior that can otherwise appear counterintuitive. Those who continue to use the construct argue that it retains practical value despite its limited scientific foundations. Strentz (1980) makes the most forceful version of this case, arguing that negotiators should actively cultivate the bond, since a captor who comes to see a hostage as a person may become less willing to harm that person. McKenzie (2004) grants that traditional accounts lack an empirical basis but suggests that, when viewed through experimental social psychology, Stockholm syndrome can still provide hostage negotiators with a basis for prediction. De Fabrique et al. (2007), working from the FBI's Hostage Barricade Database System, similarly treat the phenomenon as a guide negotiators can act on. These arguments establish the value of the label to practitioners, not its standing as a distinct clinical entity. A heuristic that helps a negotiator anticipate behavior can be serviceable even if no bounded syndrome underlies it, and McKenzie's redefinition locates the construct's practical value in general social-psychological processes rather than in the traditional syndrome account.

Despite its theoretical weaknesses, those arguing for its continued use hold that the construct describes something observable in those who have endured captivity. Speckhard et al. (2005) rest the construct's continued use on direct observation, interviewing eleven survivors of the 2002 Moscow theater siege and finding what they term Stockholm effects in all but one. They treat this response as protective and as a reaction clinicians and those preparing hostages should expect and account for. This is the most direct form of the continued-use claim: that the label names a recurring reaction worth recognizing in practice, and the study earns access to it through the victims' own accounts rather than press reports. Its limit is that documenting the reaction and finding it clinically useful establish that the responses occur, not that they form a syndrome. The reactions the authors describe are consistent with the appeasement response set out by Cantor and Price (2007) and Bailey et al. (2023), and the ten cases in which effects were reported do not clearly distinguish a discrete condition from a general response to inescapable threat.

A further argument for continued use holds that the construct can be assessed as a distinct condition through dedicated measurement scales, advancing the case that what the scales capture is a real and bounded phenomenon. As the analysis of clinical validity above shows, the most widely used such scale has not held a stable structure across samples and appears to register general distress rather than anything specific to a captor bond, and the measurement claim has nonetheless persisted, with one recent study treating Stockholm syndrome as a measurable outcome and reporting that technology-mediated partner violence contributed to it more than in-person violence (Hertlein & Teichert, 2023).

Across these arguments, the strongest cases for retaining Stockholm syndrome share a common shortfall. Each shows that the behavior is real, that the label is useful to those who work with it, and that responses associated with it can be elicited and scored, yet none establishes Stockholm syndrome as a distinct diagnostic entity, and the mechanisms these authors invoke account for the behavior without it. Justifying its continued use as a distinct syndrome would require a more sustained program of validation than the literature has yet produced.

Alternative Frameworks

While other theories, including coercive control, betrayal trauma, learned helplessness, response-based practice, and complex PTSD, address some aspects of the behaviors alleged in Stockholm syndrome (see **Table 1**), the construct's central problem is that it conflates two distinct phenomena that require separate explanations. Acute survival behavior during captivity and the durable psychological attachments that develop through prolonged cycles of abuse are not the same process, do not share the same mechanism, and should not be explained by the same term. Appeasement and trauma bonding, taken together, provide that explanation with greater precision and without pathologizing the people involved.

Cantor and Price (2007) first proposed appeasement as a natural mammalian defense response to entrapment, drawing on evolutionary theory and cross-species comparisons to argue that behaviors attributed to Stockholm syndrome reflect pacification and submission responses that serve a de-escalating function in dangerous situations. Bailey et al. (2023) extended this framework through Polyvagal Theory, describing the reflexive neurobiological process by which captives appear emotionally connected to their captors during acute threat. When a person faces a life threat, foundational survival circuits originating in the brainstem take over, moving the nervous system into a defensive state that supplants intentional behavior and displaces authentic sociality. What results is not bonding in any relational sense but a co-regulatory behavior aimed at calming the threat source long enough to survive. The motivation is not affection but safety, and the process is not psychological dysfunction but an evolved mammalian response to entrapment. Applied to the Norrmalmstorg case, this framework accounts for the hostages' documented behavior. Enmark and the other hostages were managing an armed, volatile captor over six days while police actions repeatedly escalated the danger around them. Their behavior was acute, situationally rational, and aimed at staying alive. Their criticism of police following release has been read as reasoned objection rather than symptom (Wade, 2015; Åse, 2015). Appeasement accounts for this without invoking pathology.

Trauma bonding theory addresses a different set of conditions. D. G. Dutton and Painter (1981) argued that two structural features produce durable emotional attachment, namely a marked power imbalance and intermittent abuse alternating with positive or neutral treatment. The mechanism is intermittent reinforcement, which produces stronger and more persistent behavioral responses than consistent punishment or consistent affection. This framework does not describe an acute survival response but the gradual conditioning that occurs across sustained cycles of harm and relief. It is the appropriate framework for prolonged domestic abuse, human trafficking, cult membership, and other contexts where the captive or victim remains in the relationship over time and develops loyalty that persists after the immediate threat has passed. D. G. Dutton and Painter (1993a, 1993b) tested this empirically and found that intermittency and power differentials predicted the strength of emotional attachment, locating the cause in structural dynamics rather than victim pathology. Reid et al. (2013) and Casassa et al. (2021) found consistent support for the same mechanism across a range of exploitative contexts.

The distinction matters because Stockholm syndrome has been applied indiscriminately across both types of situation, from brief hostage incidents to years of domestic abuse, with no account of why the same underlying condition would arise from such different circumstances. Appeasement explains what happens when a person co-regulates with a captor to survive immediate danger. Trauma bonding explains what happens when intermittent reinforcement across extended time produces attachment that outlasts the threat. In practice, this means the first question a clinician, law enforcement officer, or legal practitioner should ask is not whether a survivor displays the surface features attributed to Stockholm syndrome, but which conditions were actually present. Was the behavior occurring during acute threat, in which case appeasement provides the operative framework? Did it persist and deepen across sustained cycles of harm and relief, in which case trauma bonding applies? That distinction changes the intervention, changes how testimony is evaluated, and changes whether the burden of explanation falls on the victim's psychology or on the conditions that shaped their behavior. Neither framework requires a syndrome, and neither requires concluding that something is wrong with the person who survived.

Table 1.

Comparison of Stockholm Syndrome and Proposed Alternative Frameworks

Framework	Core assumptions	Behaviors explained	Most applicable contexts	Empirical support	Limitations and criticisms
Stockholm syndrome	Captive forms a paradoxical bond or identification with captor.	Defends captor; distrusts rescuers; resists testimony; appears aligned with captor.	Hostage incidents; later extended to domestic violence, trafficking, cults, and child abuse.	Weak: no DSM/ICD status, no validated criteria, mostly case reports, rare in FBI data (Fuselier, 1999; Namnyak et al., 2008).	Unstable definition; features do not reliably co-occur; pathologizes survival behavior; gendered or victim-blaming use.

Appeasement	Acute threat-management response: pacify the danger source to survive.	Placating; calming; mirroring; strategic compliance; avoiding escalation.	Hostage/captivity events, acute assault, and immediate interpersonal threat.	Supported mainly by evolutionary, clinical, and captivity-response theory (Cantor & Price, 2007; Bailey et al., 2023).	Limited direct testing; Polyvagal basis contested; explains acute behavior, not durable attachment.
Trauma bonding	Power imbalance plus intermittent abuse and relief can create durable attachment.	Staying; returning; loyalty; minimizing harm; defending abuser after immediate threat ends.	Prolonged domestic violence, trafficking, cult involvement, child abuse, and exploitation.	Empirical support for power imbalance and intermittency; reviews support use across exploitation contexts (D. G. Dutton & Painter, 1993a, 1993b; Casassa et al., 2021).	Inconsistent definitions; often conflated with Stockholm syndrome or codependency; mostly correlational evidence.
Coercive control	Abuse is a pattern of domination that restricts autonomy and liberty.	Compliance; isolation; constrained help-seeking; staying under threat; loss of autonomy.	Intimate partner violence and other dependency systems organized around control.	Developed in intimate partner violence literature, validated measures, and links to PTSD/depression (M. A. Dutton & Goodman, 2005; Lohmann et al., 2024).	Definitions and measures vary; explains control conditions better than emotional attachment.
Betrayal trauma	Harm by a trusted or depended-upon person adds a betrayal dimension beyond fear.	Dissociation; delayed disclosure; minimization; continued attachment to harmful trusted person.	Caregiver abuse, intimate partner abuse, institutional betrayal, and trusted-person exploitation.	Studies link betrayal with dissociation, memory disruption, and health symptoms (Freyd, 1998; Goldsmith et al., 2012).	Traumatic-amnesia claims disputed; evidence often self-report/correlational; less suited to brief hostage events.
Learned helplessness	Repeated uncontrollable harm lowers perceived agency and expected efficacy of action.	Passivity; hopelessness; reduced escape attempts; delayed or absent help-seeking.	Chronic abuse, captivity, institutional powerlessness, and repeated failed escape/help-seeking.	Strong experimental base on uncontrollability; influential in abuse literature (Seligman, 1975; Walker, 1980).	Can imply victim passivity; obscures resistance and real barriers; does not explain attachment to captor.
Response-based practice	Victims actively respond to violence; apparent passivity may be strategic resistance.	Strategic compliance; silence; partial disclosure; refusal; safety planning; subtle resistance.	Clinical, forensic, and advocacy interpretation of domestic violence, sexual assault, child abuse, and captivity.	Qualitative and discourse-based practice literature; useful for avoiding victim-pathologizing accounts (Coates & Wade, 2004; Wade, 2015).	Not diagnostic or causal; limited quantitative testing; requires detailed context to apply well.

Complex PTSD	Prolonged trauma can produce PTSD plus emotion, self-concept, and relationship disturbance.	Emotion dysregulation; shame; distrust; dissociation; relationship instability; difficulty feeling safe.	Chronic interpersonal trauma, childhood abuse, captivity, trafficking, torture, and prolonged violence.	Recognized in ICD-11 and supported by validated measures distinguishing PTSD/CPTSD (Cloitre et al., 2018; Herman et al., 2022).	Describes post-trauma effects, not captor-directed attachment; diagnostic boundaries remain debated.

However, both appeasement and trauma bonding carry their own limitations and remain subjects of ongoing scholarly debate. Polyvagal Theory itself has faced criticism regarding the extent of its empirical base, and the appeasement framework is recent enough that independent replication remains limited (Grossman, 2023). Trauma bonding has been criticized for inconsistent application and for failing to explain attachment across every context in which it has been invoked. Acknowledging these limitations is not a concession to Stockholm syndrome but a commitment to holding replacement frameworks to the same standard of scrutiny the article has applied throughout.

Implications Beyond the Case

The Norrmalmstorg robbery and the construct of Stockholm syndrome that emerged from it offer more than a disputed label. They provide a case of how a construct can become institutionally durable without obtaining scientific credibility. The staying power of the term came not from evidence but from wide circulation. As the term moved from media commentary to police training, from police training to courtrooms, and from courtrooms to academic and public discourse, each new site of use made the construct appear more established than it was. This paper suggests that this process represents a recognizable and replicable pattern, referred to here as circular institutional legitimation. Specifically, circular institutional legitimation is the process through which a psychological, clinical, or forensic construct acquires apparent scientific authority by way of a sequence of adoptions, citations, or uses across professional domains without any institution independently establishing its empirical validity. The construct appears credible not because it has been reliably defined, measured, replicated, or distinguished from competing explanations, but because it has already circulated through respected professional settings.

The defining feature of circular institutional legitimation is epistemic substitution. That is, the normal requirements for establishing a construct's validity are replaced by evidence of institutional uptake and cross-use. A court cites the construct because clinicians or law enforcement personnel have used it. Clinicians cite it because it appears in legal, forensic, or training contexts. Textbooks reproduce it because it has become recognizable in professional and public discourse. Each site of use treats prior institutional use as if it were proof that the construct has already been validated elsewhere. The result is a closed loop in which apparent authority accumulates without empirical grounding ever entering the chain. Recent overviews continue to present the construct as an established area of inquiry while acknowledging that it has no diagnostic criteria and no place in the classification systems (Inić, 2025; Logan, 2018).

This process has several linked features. The construct often begins outside ordinary empirical validation, for instance through media commentary, professional opinion, legal argument, or interpretation of a highly visible case. Then, it is adopted by a high-status institution or individual for practical or explanatory reasons. Subsequently, it moves across domains, for example from law enforcement to courts, from courts to clinical discourse, and from clinical discourse to textbooks and public understanding. This repeated use displaces the question of validation. The relevant question becomes not "Has this construct been demonstrated to exist?" but "Has this construct already been recognized?" As a result, responsibility for validation becomes distributed. No single institution takes responsibility for proving the construct because each assumes that another institution has already done so. This culminates in the construct becoming resistant to correction because its authority no longer depends on one source. It is dispersed across multiple professional systems.

Stockholm syndrome mirrors this pattern closely. The concept emerged from Bejerot's interpretation of the Norrmalmstorg hostages' behavior, despite his lack of direct clinical examination of the people whose reactions he described. It was then taken up by law enforcement as a practical framework for hostage negotiation, introduced into American legal consciousness through the Hearst trial, formalized by Ochberg's later account, and reproduced in clinical, forensic, academic, and popular settings. At no point in this chain did the concept undergo the ordinary process of construct validation. There were no stable diagnostic criteria, no validated measurement instruments, no controlled studies establishing a discrete syndrome, and no clear demonstration that the behaviors attributed to Stockholm syndrome were distinct from other responses to coercion, captivity, or threat. What accumulated was not evidence but institutional familiarity and proliferation of use.

Circular institutional legitimation is analogous to existing theories of knowledge production but identifies a narrower, more specific mechanism (see **Table 2**). Gieryn's (1983) concept of boundary work describes how scientific authority is constructed and defended through rhetorical and institutional demarcation rather than fixed epistemic criteria. However, circular institutional legitimation is different. It is not primarily about how science marks itself off from non-science. It is about how a construct can acquire the appearance of scientific or clinical authority when institutions repeatedly rely on one another's prior use of the construct instead of requiring direct validation. The concept also differs from Hacking's (1995) account of looping effects. Looping effects describe how classifications can affect the people classified, who may then understand themselves or behave differently in response to the classification. Circular institutional legitimation describes a different problem: how the classification itself becomes authoritative before it has been shown to be valid. In other words, looping effects concern the interaction between classifications and classified people. Circular institutional legitimation concerns the interaction among institutions that mutually reinforce the standing of a classification.

It also departs from ordinary diffusion. Many ideas spread across institutions, and diffusion alone does not make a concept invalid. Circular institutional legitimation occurs when the spread of a concept is mistaken for evidence of its validity. The closest institutional analogue is mimetic isomorphism, DiMaggio and Powell's (1983) account of how organizations facing uncertainty adopt the forms and practices of others to secure legitimacy. That process operates on organizational structure and proceeds whether or not the imitated practice is effective, so a construct's validity is never what is in question, whereas circular institutional legitimation operates specifically on empirical validity and on the construct's standing being carried from one domain to the next in place of empirical validation. The construct does not merely circulate. Its circulation becomes part of the reason it is treated as legitimate. Professional familiarity can easily be confused with empirical support, especially when a term is useful, memorable, and already embedded in applied practice. Therefore, circular institutional legitimation also helps explain why weak constructs become difficult to dislodge once embedded. By the time concerns about validity emerge, the concept is no longer housed in one location. Instead, it is distributed across professions. Adorjan et al.'s (2012) description of Stockholm syndrome as a vernacular resource helps clarify the later stages of this process. Their analysis shows how the label moved beyond hostage situations and became available as a rhetorical device in debates about domestic violence, human trafficking, race, politics, and international relations. Circular institutional legitimation explains how the concept became available for that wider use in the first place. Once institutional circulation had given Stockholm syndrome the appearance of legitimacy, the term could be redeployed in new settings even when those settings differed substantially from the original hostage context.

Table 2.
Comparison of Circular Institutional Legitimation and Related Concepts

Concept	Main focus	Mechanism	Difference from circular institutional legitimation
Circular institutional legitimation	How weak constructs gain apparent scientific standing.	Repeated adoption across institutions substitutes for empirical validation.	Anchor concept. Its defining feature is epistemic substitution, with institutional uptake treated as though

			validation had occurred elsewhere.
Mimetic isomorphism (DiMaggio & Powell, 1983)	How organizations come to resemble one another.	Organizations facing uncertainty imitate others they regard as legitimate or successful.	Concerns organizational form and legitimacy, not whether a construct was ever validated.
Ordinary diffusion	How ideas spread across settings.	A concept moves from one field or institution to another.	Describes spread as such, whereas this account is the case where spread is mistaken for evidence.
Vernacular resource (Adorjan et al., 2012)	How a term becomes available for public argument.	A label is used rhetorically across social and political settings.	Explains later public reuse once the term already circulates, not how it first acquired institutional standing.
Boundary work (Gieryn, 1983)	How science distinguishes itself from non-science.	Institutions and experts draw boundaries around what counts as legitimate knowledge.	Concerns demarcation, not whether a given construct was ever validated.
Looping effects (Hacking, 1995)	How classifications affect the people classified.	Classifying people changes them, and those changes feed back so the category itself must be revised.	Concerns the interaction between a label and the people it labels, not how institutions confer the label's authority.

The value of circular institutional legitimation lies in identifying constructs that hold their standing through institutional circulation, not evidence. It has significant consequences for clinical judgment, legal proceedings, forensic evaluation, and public interpretation of victims' behavior, where a label can shape credibility, diagnosis, and testimony before any validation has occurred. The concept does not imply that the institutions adopting such a construct act in bad faith, nor that the behaviors associated with it are unreal. Captives may cooperate with their captors and grow hostile toward the authorities trying to help them. The question is whether those behaviors constitute a distinct syndrome. Circular institutional legitimation explains how a label can stay in professional use even when that question has never been answered.

Part of why the term persists is that it proved useful. Hostage negotiators could anticipate a captive's hostility toward police. Courts had a way to interpret coerced conduct. Clinicians had a name for the bonds that sometimes form under captivity and abuse. But usefulness is not validity. A construct can help practitioners organize what they observe while still failing to identify a distinct psychological condition. The Norrmalmstorg case suggests that weak constructs should be judged not only by their descriptive value but by how they acquired their authority. A construct established through empirical research and replication can be argued over on empirical grounds. A construct established through circulation and applied use presents a more basic problem.

The same pattern reaches past this single case. Any psychological or forensic concept that holds its standing mainly because other fields or institutions already use it is vulnerable to the same kind of entrenchment. This is not an argument for rejecting such concepts automatically. It is an argument for examining them more closely. Does the construct have stable criteria? Can it be measured reliably? Does it predict anything the existing explanations do not? Has it been replicated across settings? Does its institutional history show empirical validation or only repeated adoption? Where the route to authority has been circular and not empirical, correction will take more than better research. It will take coordinated skepticism across the institutions that produced the appearance of legitimacy.

Limitations

This review has several limitations that should be acknowledged when interpreting the findings. Since this is a narrative review rather than a systematic review, source selection reflects the author's judgment. A defined database search and deliberate review of scholarship defending the construct were used to reduce selection bias, although residual selection bias remains possible. The empirical base is thin and weighted toward case reports, so the conclusions rest on a limited record, and only a prospective study using validated measures could settle the validity question. Much of the primary material on the case is retrospective and in Swedish and accessed through translation, so some details may have shifted, and a few remain disputed despite checks against the closest available sources. The accounts weighted most heavily, Enmark's and Olofsson's, also reject the syndrome, so on that point the source weighting and the conclusion overlap. Finally, the reading of the concept's history as a self-reinforcing loop is an interpretation, not a demonstrated finding, built on one case, so it holds only until the same pattern is tested across other contested constructs.

Conclusion

Stockholm syndrome has held a fixed place in clinical training, forensic testimony, and public understanding for over fifty years without earning that place. The concept appears in no recognized diagnostic system, has never been tested through controlled empirical study, and was coined by a psychiatrist who did not examine the people he claimed to diagnose. That none of this has dislodged the concept from professional use tells us something about how psychological knowledge is maintained. The distance between what the concept has been permitted to do in professional settings and what the evidence actually supports is not an oversight waiting to be corrected. This paper has argued that it is the product of a process it terms circular institutional legitimation, in which each adopting institution treated the prior use of the concept by other institutions as a substitute for the empirical validation that never occurred. Stockholm syndrome is almost certainly not the only construct to have traveled this route. Any psychological concept whose standing depends on the fact that other fields already use it rather than on whether anyone has demonstrated its validity is vulnerable to the same kind of entrenchment, and the professions that adopted it without scrutiny bear the responsibility for subjecting it to the scrutiny it was never given. Future scholarship should continue to interrogate how psychological concepts achieve cultural authority independent of their evidentiary base and attend more carefully to the voices of those who have been diagnosed without consent, examined without contact, and dismissed without cause.

References

Adorjan, M., Christensen, T., Kelly, B., & Pawluch, D. (2012). Stockholm syndrome as vernacular resource. *The Sociological Quarterly*, *53*(3), 454–474. https://doi.org/10.1111/j.1533-8525.2012.01241.x

Ahmad, A., Aziz, M., Anjum, G., & Mir, F. V. (2018). Intimate partner violence and psychological distress: Mediating role of Stockholm syndrome. *Pakistan Journal of Psychological Research*, *33*(2), 541–557.

Alexander, S. (1979). Anyone's daughter: The Times and trials of Patty Hearst. The Viking Press.

Alshwaiheen, N., Saleh, D., & Alani, W. (2025). A case of post-traumatic stress disorder (PTSD) complicated by Stockholm syndrome: A unique psychiatric phenomenon in the context of intimate partner violence. *Cureus*, *17*(4). https://doi.org/10.7759/cureus.82307

Aponte, R. (2020). An interview with Frank Ochberg on PTSD. Psychotherapy.net. https://www.psychotherapy.net/perspectives/articles/frank-ochberg-on-treating-ptsd/

Appelbaum, P. S. (2017). Reflections on the Goldwater Rule. *The Journal of the American Academy of Psychiatry and the Law*, *45*(2), 228–232.

Åse, C. (2015). Crisis narratives and masculinist protection. *International Feminist Journal of Politics*, *17*(4), 595–610. https://doi.org/10.1080/14616742.2015.1042296

Bailey, R., Dugard, J., Smith, S. F., & Porges, S. W. (2023). Appeasement: Replacing Stockholm syndrome as a definition of a survival strategy. *European Journal of Psychotraumatology*, *14*(1). https://doi.org/10.1080/20008066.2022.2161038

Bejerot, N. (1974). The six day war in Stockholm. *New Scientist*, *61*(886), 486–487.

Binder, L. M., & Thompson, L. L. (1995). The ethics code and neuropsychological assessment practices. *Archives of Clinical Neuropsychology*, *10*(1), 27–46. https://doi.org/10.1016/0887-6177(93)e0004-u

Cantor, C., & Price, J. (2007). Traumatic entrapment, appeasement and complex post-traumatic stress disorder: Evolutionary perspectives of hostage reactions, domestic abuse and the Stockholm syndrome. *Australian & New Zealand Journal of Psychiatry*, *41*(5), 377–384. https://doi.org/10.1080/00048670701261178

Casassa, K., Knight, L., & Mengo, C. (2021). Trauma bonding perspectives from service providers and survivors of sex trafficking: A scoping review. *Trauma, Violence, & Abuse*, *23*(3), 969–984. https://doi.org/10.1177/1524838020985542

Cerny, C. A., Friedman, S. H., & Knoll, J. L. (2010). Stockholm Syndrome. Correctional Mental Health Report, 12(2), 17–32.

Cloitre, M., Shevlin, M., Brewin, C. R., Bisson, J. I., Roberts, N. P., Maercker, A., Karatzias, T., & Hyland, P. (2018). The International Trauma Questionnaire: Development of a self‐report measure of ICD‐11 PTSD and complex PTSD. *Acta Psychiatrica Scandinavica*, *138*(6), 536–546. https://doi.org/10.1111/acps.12956

Coates, L., & Wade, A. (2004). Telling it like it isn't: Obscuring perpetrator responsibility for violent crime. *Discourse & Society*, *15*(5), 499–526. https://doi.org/10.1177/0957926504045031

De Fabrique, N., Romano, S. J., Vecchi, G. M., & Van Hasselt, V. B. (2007). Understanding Stockholm syndrome. *FBI Law Enforcement Bulletin*, *76*(7), 10–15.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, 48(2), 147–160. https://doi.org/10.2307/2095101

Dutton, D. G., & Painter, S. (1993a). Emotional attachments in abusive relationships: A test of traumatic bonding theory. *Violence and Victims*, *8*(2), 105–120. https://doi.org/10.1891/0886-6708.8.2.105

Dutton, D. G., & Painter, S. (1993b). The battered woman syndrome: Effects of severity and intermittency of abuse. *American Journal of Orthopsychiatry*, *63*(4), 614–622. https://doi.org/10.1037/h0079474

Dutton, D. G., & Painter, S. L. (1981). Traumatic bonding: The development of emotional attachments in battered women and other relationships of intermittent abuse. *Victimology: An International Journal*, *6*(1), 139–155.

Dutton, M. A., & Goodman, L. A. (2005). Coercion in intimate partner violence: Toward a new conceptualization. *Sex Roles*, *52*(11–12), 743–756. https://doi.org/10.1007/s11199-005-4196-6

Enmark, K., & Wesslén, G. (2020). *Jag blev Stockholmssyndromet*. SAGA Egmont.

Ewing, C. P., & McCann, J. T. (2006). Minds on trial: Great cases in law and psychology. Oxford University Press.

Freyd, J. J. (1998). *Betrayal trauma: The logic of forgetting childhood abuse*. Harvard University Press.

Fuselier, G. (1999). Placing the Stockholm syndrome in perspective. *FBI Law Enforcement Bulletin*, *68*(7), 22–25. https://doi.org/10.1037/e313152004-006

Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, *48*(6), 781. https://doi.org/10.2307/2095325

Goldsmith, R. E., Freyd, J. J., & DePrince, A. P. (2012). Betrayal trauma. *Journal of Interpersonal Violence*, *27*(3), 547–567. https://doi.org/10.1177/0886260511421672

Graham, D. L. R., Rawlings, E. I., & Rigsby, R. K. (1994). *Loving to survive: Sexual terror, men's violence, and women's lives*. New York University Press.

Graham, D. L., Rawlings, E. I., Ihms, K., Latimer, D., Foliano, J., Thompson, A., Suttman, K., Farrington, M., & Hacker, R. (1995). A scale for identifying "Stockholm syndrome" reactions in young dating women: Factor structure, reliability, and validity. *Violence and Victims*, *10*(1), 3–22. https://doi.org/10.1891/0886-6708.10.1.3

Grossman, P. (2023). Fundamental challenges and likely refutations of the five basic premises of the polyvagal theory. *Biological Psychology*, *180*, 108589. https://doi.org/10.1016/j.biopsycho.2023.108589

Grubb, A. R. (2023). Police hostage and crisis negotiation training: Foci, protocols, and best practice principles. *Police conflict management, volume II*, 293–324. https://doi.org/10.1007/978-3-031-41100-7_13

Hacker, F. (1977). Crusaders, criminals, crazies: Terror and terrorism in our time. W. W. Norton & Company.

Hacking, I. (1995). *Rewriting the soul: Multiple personality and the sciences of memory*. Princeton University Press.

Herman, J. L., Mathews, A., & Sands, X. (2022). *Trauma and recovery: The aftermath of violence — from domestic abuse to political terror*. Basic Books.

Hertlein, K. M., & Teichert, R. (2023). The association between technology-mediated intimate partner violence and Stockholm syndrome. *Journal of Couple & Relationship Therapy*, *22*(3), 227–241. https://doi.org/10.1080/15332691.2023.2165583

Hill, J. (2020). *See what you made me do: The dangers of domestic abuse that we ignore, explain away, or refuse to see*. Sourcebooks.

Inić, T. (2025). Stockholm syndrome: A dimension of trauma. *Sanamed*, *20*(1), 97–102. https://doi.org/10.5937/sanamed0-57254

Jameson, C. (2010). The "short step" from love to hypnosis: A reconsideration of the Stockholm syndrome. *Journal for Cultural Research*, *14*(4), 337–355. https://doi.org/10.1080/14797581003765309

Justia. (2018, May 29). *People v Abdur-Razzaq*. https://law.justia.com/cases/new-york/other-courts/2018/2018-ny-slip-op-28161.html

Jülich, S. (2005). Stockholm syndrome and child sexual abuse. *Journal of Child Sexual Abuse*, *14*(3), 107–129. https://doi.org/10.1300/j070v14n03_06

King, D. (2020). *Six days in August: The story of Stockholm syndrome*. W. W. Norton & Company.

Kroll, J., & Pouncey, C. (2016). The ethics of APA's Goldwater Rule. *The Journal of the American Academy of Psychiatry and the Law*, *44*(2), 226–235.

Lang, D. (1974, November 18). *The bank drama*. The New Yorker. https://www.newyorker.com/magazine/1974/11/25/the-bank-drama

Logan, M. H. (2018). Stockholm syndrome: Held hostage by the one you love. *Violence and Gender*, *5*(2), 67–69. https://doi.org/10.1089/vio.2017.0076

Lohmann, S., Cowlishaw, S., Ney, L., O'Donnell, M., & Felmingham, K. (2024). The trauma and mental health impacts of coercive control: A systematic review and meta-analysis. *Trauma, Violence, & Abuse*, *25*(1), 630–647. https://doi.org/10.1177/15248380231162972

McKenzie, I. K. (2004). The Stockholm syndrome revisited. *Journal of Police Crisis Negotiations*, *4*(1), 5–21. https://doi.org/10.1300/j173v04n01_02

Mouafo, A., & Lontio, S. (2024). Stockholm syndrome and gender-related ideologies and attitudes: A psychometric assessment. *Psychology and Behavioral Sciences*, *13*(3), 56–74. https://doi.org/10.11648/j.pbs.20241303.11

Namnyak, M., Tufton, N., Szekely, R., Toal, M., Worboys, S., & Sampson, E. L. (2008). 'Stockholm syndrome': Psychiatric diagnosis or urban myth? *Acta Psychiatrica Scandinavica*, *117*(1), 4–11. https://doi.org/10.1111/j.1600-0447.2007.01112.x

Obeid, S., & Hallit, S. (2018). Correlation of the Stockholm syndrome and early maladaptive schemas among Lebanese women victims of beating into domestic/marital violence. *Couple and Family Psychology: Research and Practice*, *7*(3–4), 171–182. https://doi.org/10.1037/cfp0000106

Ochberg, F. M., & Soskis, D. A. (1982). *Victims of terrorism - Westview special studies in national and international terrorism*. Westview Press.

Olofsson, C., & Grimwalker, L. (2023). *Skit på dej: En rövarhistoria*. Mondial.

Rahaim, A., & Matusitz, J. (2020). Patty Hearst and the Symbionese Liberation Army (SLA): An examination through differential association theory. Journal of Applied Security Research, 15(3), 408–422. https://doi.org/10.1080/19361610.2019.1710094

Rahme, C., Haddad, C., Akel, M., Khoury, C., Obeid, H., Obeid, S., & Hallit, S. (2020). Does Stockholm syndrome exist in Lebanon? Results of a cross-sectional study considering the factors associated with violence against women in a Lebanese representative sample. *Journal of Interpersonal Violence*, *36*(23–24), 11509–11531. https://doi.org/10.1177/0886260519897337

Reid, J. A., Haskell, R. A., Dillahunt-Aspillaga, C., & Thor, J. A. (2013). Contemporary review of empirical and clinical studies of trauma bonding in violent or exploitative relationships. *International Journal of Psychology Research*, *8*(1), 37–73.

Rizo-Martínez, L. E. (2018). El síndrome de Estocolmo: Una revisión sistemática. *Clínica y Salud*, *29*(2), 81–88. https://doi.org/10.5093/clysa2018a12

Santoyo Telles, F., & Rizo Martínez, L. E. (2019). Spanish translation and validation of the scale for identifying "Stockholm syndrome" reactions (SISSR) related to dating violence in Mexican women. *Revista Argentina de Clínica Psicológica*, *28*(5), 849–858. https://doi.org/10.24205/03276716.2019.1110

Schlossberg, H., & Freeman, L. (1974). *Psychologist with a gun*. Coward, McCann & Geoghegan.

Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death (a series of books in psychology)*. W. H. Freeman.

Speckhard, A., Tarabrina, N., Krasnov, V., & Mufel, N. (2005). Stockholm effects and psychological responses to captivity in hostages held by suicide terrorists. *Traumatology*, *11*(2), 121–140. https://doi.org/10.1177/153476560501100206

Strentz, T. (1979). Law enforcement policy and ego defenses of the hostage. *FBI Law Enforcement Bulletin*, *48*(4), 2–12.

Strentz, T. (1980). The Stockholm syndrome: Law enforcement policy and ego defenses of the hostage. *Annals of the New York Academy of Sciences*, *347*(1), 137–150. https://doi.org/10.1111/j.1749-6632.1980.tb21263.x

Stricker, G. (2010). American Psychological Association code of ethics. *The Corsini Encyclopedia of Psychology*, 1–2. https://doi.org/10.1002/9780470479216.corpsy0047

Taylor, S. E. (2012). Tend and befriend theory. *Handbook of theories of social psychology: Volume 1*, *1*, 32–49. https://doi.org/10.4135/9781446249215.n3

U.S. Department of State. (2020, June). *Trauma bonding in human trafficking*. https://www.state.gov/wp-content/uploads/2020/10/TIP_Factsheet-Trauma-Bonding-in-Human-Trafficking-508.pdf

Wade, A. (2015). Dignity 2015: Hawkes Bay, Aotearoa. In *The myth of "Stockholm syndrome" (and other concepts invented to discredit women victims of violence)*. Hawke's Bay, New Zealand.

Walker, L. E. (1980). *Battered woman: An authoritative and compassionate guide to escaping abuse and reclaiming your life*. Harper and Row.

West, L. J., & Martin, P. R. (1996). Pseudo-identity and the treatment of personality change in victims of captivity and cults. *Cultic Studies Journal*, *13*(2), 125–152.

Zhang, T., & Datta, V. (2022). Expert testimony in sex trafficking cases. *Journal of the American Academy of Psychiatry and the Law*, *50*(2), 1–9. https://doi.org/10.29158/JAAPL.210051-21

The View from Inside: Ranked Choice Voting and New York City

Anita Chadha
University of Houston, Downtown

Abstract
New Yorkers have used ranked choice voting (RCV) for their past three primary elections. Each iteration of their election has shown New Yorkers' understanding and appreciation of RCV processes. In addition, their last RCV election in 2025 represented the largest use to date of ranked choice voting in a citywide election in the U. S. During this election, voters ranked candidates for office in 38 contests across the city, including the competitive Democratic primary. Voters embraced the method with a vast majority of New Yorkers ranking at least two candidates, one of which resulted in more representative outcomes. Of particular importance was their mayoral race, where Zohran Mamdani defeated Andrew Cuomo 56 percent to 44 percent, becoming the city's first Muslim mayor. In addition, their 2025 RCV election was particularly significant as one million New Yorkers used the method with 96% of voters finding the ballot simple to complete and more than 75% wanting to expand RCV use in future elections.

Keywords: Rank choice vote, New York, Electoral reform

Introduction

Ranked-choice voting (RCV) is an electoral reform in which voters rank candidates in order of preference: first, second, third, and so on. Voters can rank as many or as few candidates as they like, as ranking other candidates does not affect their first choice. However, ranking provides backup choices if their favorite candidate does not win. The counting of ballots commences with tallying voters' first choices. If a candidate receives more than half of the first choices, that candidate wins, just like in any other election. If there is no majority winner after counting first choices, the race is decided by an “instant runoff.” The candidate with the fewest votes is eliminated, and voters who picked that candidate as their top choice will have their next choice counted. This process continues until there’s a majority winner, a candidate who wins with more than half of the vote.

RCV, 2021

New York's journey with RCV began in 2019 when a New York City charter commission voted 12-0 to place a charter amendment on that year’s ballot that would establish RCV for all primary and special elections (FairVote, 2019; Chadha, 2019). A large majority of New Yorkers (73%) passed RCV that year, and quickly the city began its educational campaign for voters, poll workers, election judges, and candidates, as

RCV necessitated voter education on ranking processes as well as for candidates' strategies as a candidate could win without being a voter's first choice (FairVote, 2021).

RCV, Education: Ballot and how to rank.

The city created a website specific to RCV, along with images of what the RCV ballot would look like and explained the ranking processes. In addition, online voter guides and interactive tools were provided. Various other electoral components had to be addressed, such as the ballot design and a space for voters to practice how to correctly rank, as a correctly marked ballot was one where the voter has indicated a first, second, and third choice, as shown in Figure 1.

Figure 1: How to correctly rank in RCV.

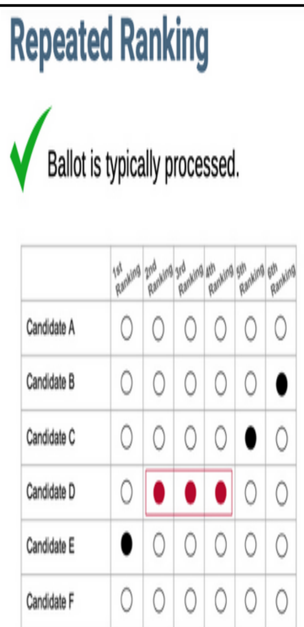
Source: Retrieved from https://fairvote.org/resources/ranked-choice-voting-how-tos/

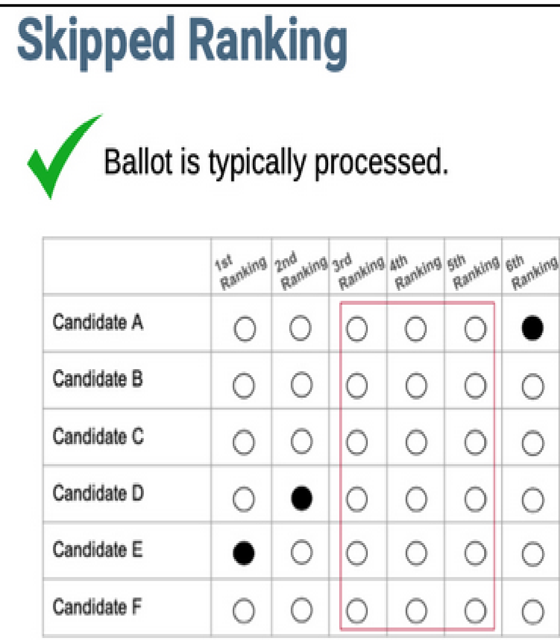
In addition, education, and awareness of incorrect ways to rank were also provided. For instance, an incorrectly marked ballot would be if they 'repeated' their ranking, which occurred when a voter marked the same candidate across multiple ranking positions, as shown in Table 1. In this scenario, the highest-ranked ballot was counted. The ballot was treated as though candidate E was first, D was second, C was third, and B was fourth. As a second way to incorrectly vote would be that of 'skipping' their ranking, which occurred when a voter does not rank all the available candidates and ranks more candidates at subsequent rankings below another candidate with no preferences in between. As shown in the same table 1, the ballot was counted as though the voter ranked B first, D second, and A third. A third way to incorrectly rank would be 'overvoting,' which occurs when a voter marks multiple candidates at the same ranking position. In this scenario, the ballot would count as intended for the voter's first and second choices, and as the ballot becomes impossible to interpret when the voter's first and second choices were eliminated, the remainder of the ballot preferences would be rejected, as shown in Table 1 as well.

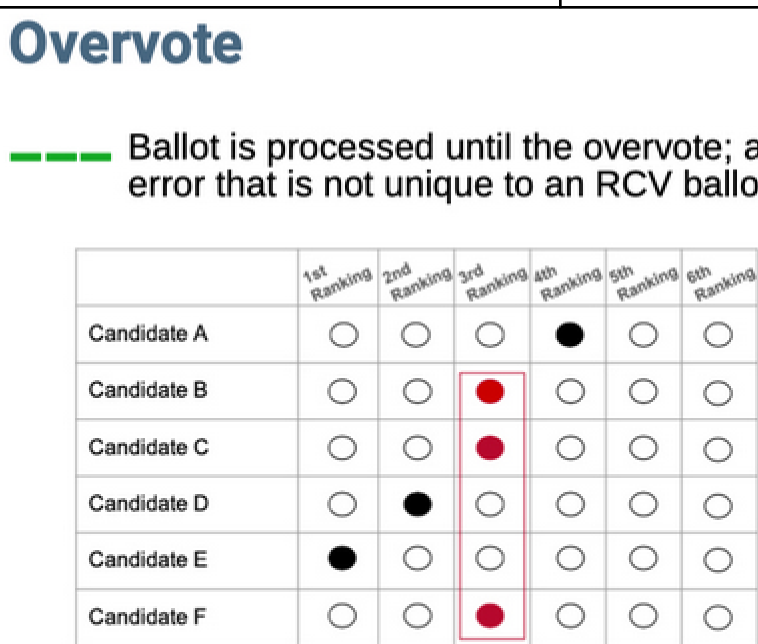
In addition, door-to-door outreach campaigns were initiated alongside leaflets and videos in thirteen languages (Arabic, Bangla, Chinese, English, French, Haitian Creole, Khmer, Korean, Nepali, Russian, Spanish, Vietnamese) (Rank the vote, N.Y, 2021).

Table 1

Incorrect rankings: Repeated, skipped and overvotes.







Post analysis, 2021

Post-election analysis of their first RCV (2021) election showed that more than 941,000 New Yorkers participated in the primaries. This was the city's highest turnout since 1989, and a large majority (92 %) of voters who voted for mayor also participated in elections for their comptroller and public advocate (FairVote, 2021). The polls found that an overwhelming majority (95 %) of New Yorkers across both racial and ethnic groups (Black, Hispanic, Asian, and white voters) found their ballot simple to complete as shown in Table 2 (Solomon, 2021). While an additional vast majority (78%) found the RCV ballot simple to complete (Rank the Vote NY (2021). Not only did the vast majority of New Yorkers understand the processes, but the election resulted in candidates that reflected the diversity of the city.

Table 2

RCV ballot simple to complete across racial groups

Black	93%
Hispanic	95%
Asian	97%
White	95%

Source: Rank the voter NYC releases Edison research exit poll on the election. Retrieved from http://readme.readmedia.com/RANK-THE-VOTE-NYC-RELEASES-EDISON-RESEARCH-EXIT-POLL-ON-THE-ELECTION/17989282?utm_source=newswire&utm_medium=email&utm_campaign=med

Women gained representation: Voters stepped up making their voices heard as they elected a diversity of candidates to office. Women who were 52% of the city's population and held only 28% of the seats on the New York City council won in large numbers. In fact, for the first time in history, women won by a majority (61%). Thus, for the first time, women held the majority in the New York City Council (Represent women, 2022).

Second-ever Black mayor: In the race for New York's mayor there were 13 candidates. In this race Eric Adams won the Democratic primary after appealing to the political center and promising to strike the right balance between fighting crime and ending racial injustice in policing. In the end, New Yorkers again elected a diverse candidate, with Adams becoming the city's second-ever Black mayor.

Campaign strategies: Moreover, as anticipated, RCV most visibly challenged campaign strategies in the race for mayor, which featured thirteen candidates. Adams' closest rivals in the race included former city sanitation commissioner Kathryn Garcia, who campaigned as a technocrat and proven problem-solver, and former City Hall legal advisor Maya Wiley, who had been endorsed by U.S. Rep. Alexandria Ocasio-Cortez. Nonetheless, after eight rounds of counting, Eric Adams won the race, becoming the city's second-ever Black mayor elected with majority consensus as shown in Table 3.

Table 3

Round-by-Round Results, NYC Democratic Mayoral Primary

Candidate	Round one	Round two	Round three	Round four	Round five	Round six	Round seven	Round eight
Eric Adams	30.7%	30.8%	30.8%	31.2%	31.7%	34.6%	40.5%	50.4%
Maya Wiley	21.4%	21.4%	21.4%	22%	22.4%	26.1%	29.1%	
Kathryn Garcia	19.6%	19.6%	19.6%	19.9%	20.5%	24.4%	30.5%	49.6%
Andrew Yang	12.2%	12.2%	12.3%	12.6%	13%	14.8%		
Scott Stringer	5.5%	5.5%	5.5%	5.7%	6.1%			
Dianne Morales	2.8%	2.8%	2.8%	3.2%	3.3%			
Raymond McGuire	2.7%	2.7%	2.7%	2.8%	3.0%			
Shaun Donovan	2.5%	2.5%	2.5%	2.6%				

Source: Lamendola, C et al., 2022: Why Women Won in 2021: How a Twin-Track Approach Advanced Women's Representation on the New York City Council

Ranking: A vast majority (83%) of New Yorkers ranked at least two candidates, while another large majority (42%) of voters maximized their newfound ranking privileges and ranked five candidates. Likewise, an overwhelming majority (95%) of voters found their ballot simple to complete, with another overwhelming majority (78%) of New Yorkers saying they understood RCV extremely or very well (Common Cause, 2023).

RCV, 2023

New York held its second RCV election in 2023, a quieter, uneventful cycle, as voters had already experienced the RCV process in 2021. Moreover, this election cycle had fewer competitive races in which there were just fourteen primaries with three or more candidates running in comparison to their 2021 RCV election.

RCV Re-education

Nonetheless, like their first RCV election, educational campaigns were undertaken by NYC Votes, a nonpartisan voter engagement initiative of the New York City Campaign Finance Board, as they reached out to the diverse communities throughout New York City (Chadha, 2019; FairVote, 2023). They printed and mailed voter guides in thirteen languages spoken in the city, which contained extensive information on how to complete the ballot, how ballots were counted, as well as providing tear-out sheets so that voters could rank their choices for each office and bring them to vote.

Education: The 2023 RCV education effort was supported by a $500,000 advertising campaign with advertisements running on popular platforms like Google, Meta (Facebook/Instagram), YouTube, Snapchat, and Twitch, as well as print, connected RCV, radio, and out-of-home placements like bus shelters and subway kiosks (Otis, 2024). Educational videos were published on YouTube and embedded on a multilingual website. A social media toolkit offering suggested captions and graphics illustrating the RCV process was translated and shared with numerous government agencies and partner organizations. Additionally, RCV presentations were conducted in Bengali, Chinese (simplified and traditional), Punjabi, Spanish, and Urdu. Yet, unlike their first RCV election in 2023, a paid "train-the-trainer" program was established, which prepared and compensated civic organizations and associations for training their own communities, further expanding RCV education to a wider audience.

Post analysis, 2023

Post-election analysis of their second RCV election reiterated their first use in that New Yorkers not only understood the method, but they also turned out in high numbers as they continued to take advantage of the opportunity to rank candidates, and that candidates continued to cross-endorse each other as a campaign strategy.

Rankings: Analogous to the 2021 post-election analysis, a majority, (63%) of voters ranked three or more candidates (Otis, 2024). Similarly, a large majority of voters understood the ranking process as ranking of candidates was comparable in both 2021 and 2023 RCV elections. Furthermore, the median error rate in both the 2021 and 2023 elections was comparable at 5 % highlighting their understanding and appreciation of ranking (Otis, 2024).

Candidate strategies: As past evidence suggested, and similar to their first RCV election, voters followed cues from candidates about how to rank. Such was the case in District 9, opponents Yusef Salaam and Al Taylor supported each other, encouraging voters to rank each other first and second. As Salaam described it, he and Taylor are "together in so many things." (Otis, 2024). In the end, there was only partial crossover voting between Salaam and Taylor despite their cross-endorsement as 41% of Salaam voters chose Taylor as their second choice, and 35% of Taylor voters ranked Salaam second, as shown in Table 4. Inez Dickens also earned strong second-choice support from supporters of both finalists.

Table 4

Crossover voting/endorsements

First choice Candidate	First choice Candidate	First choice Candidate	First choice Candidate
First choice Salaam		41%	35%
First choice Taylor	35%		36%

First choice Dickens	30%	32%	

Source: Otis, D. (2024). What we learned from New York City's second ranked choice voting election. Retrieved from https://fairvote.org/report/rcv-in-nyc-report-2023/?section=a-quieter-election-cycle

RCV round by round counting: Yet again, as past evidence suggested, RCV encouraged collaboration and opportunities for coalition-building and cross-endorsements, which allowed candidates to support each other without harming their own prospects. Such was the case again in 2023, as it was in 2021, during the ninth council district primary between opponents Yusef Salaam and Al Taylor, who not only expressed support for each other but encouraged voters to rank each other first and second. In the end, Yusef Salaam won the race with 7,058 votes in the final fourth round. Yusef Salaam defeated Inez E. Dickens, who received 3,996 final votes as shown in Table 5 (Board of Elections, 2023).

Table 5

How the votes were counted - round by round

	Round 1		Round 2			Round 3			Round 4		
Eliminated in this round			Write-ins						Kristin R. Jordan; Al Taylor[**]		
Elected						Yusef Salaam over 50%			Yusef Salaam[*]		
Votes in this round	# votes	% of vote	votes transferred	# votes	% of vote	votes transferred	# votes	% of vote	votes transferred	# votes	% of vote
Yusef Salaam	5,761	49.9%	21	5,782	50.2%		5,782	50.2%	1276	7,058	63.9 %
Inez E. Dickens	2,898	25.1%	13	2,911	25.3%		2,911	25.3%	1085	3,996	36.1 %
Al Taylor	1,664	14.4%	40	1,704	14.8%		1,704	14.8%	-XXX		
Kristin R. Jordan	1,098	9.5%	16	1,114	9.7%		1,114	9.7%	-XXX		
Write-ins	123	1.1%	-XXX								
Inactive ballots			33	33			33		457	490	

Source: Board of elections in the City of New York 2023 primary election official ranked choice rounds. Retrieved from https://www.vote.nyc/sites/default/files/pdf/election_results/2023/20230627Primary%20Election/rcv/025756_1.html

The New York City RCV primary elections in 2025 were the city's third time using RCV to elect its mayor, city council, and other local offices. This third iteration of RCV represented the largest use to date of ranked choice voting in a citywide election in the U. S. During this election cycle, voters had the opportunity to rank candidates for office in thirty eight contests across the city, including the competitive Democratic mayoral primary. Post analysis of this election found that New Yorkers used RCV to its fullest potential across several measures, including turnout, ease of use, ranking candidates, and candidate strategies, with the election resulting in candidates that reflected the diversity of the city.

Post analysis, 2025

<u>Turnout:</u> Turnout in 2025 was extraordinarily high, with a large majority of voters (79.2%) ranking their choices for at least one office on the ballot, while others ranked five candidates. This percentage was more than those who had ranked all candidates in the previously held 2021 and 2023 RCV elections. In fact, these primaries saw the highest turnout since 1989 (Hutchinson & Mohamud, 2025). Furthermore, in contests that went to an RCV count, a total of two hundred and forty three ballots ranked non-frontrunner first but counted for one of the finalist candidates (Hutchinson, 2025).

<u>Ease of use:</u> Moreover, a large majority of New Yorkers (96%) found the ballot easy to complete (Hutchinson, 2025; SurveyUSA, 2025). While New Yorkers across each ethnic group found their ballots simple to complete. In this regard, (95%) of Black voters found their ballot simple to complete, (97%) of Hispanic voters and (94%) of Asian voters found the same (Common cause, 2025). In fact, "relatively few ballots in RCV elections contained errors, and even fewer ballots were rejected, showcasing voter understanding of the method (Otis, 2025).

<u>Ranking</u> post-election analysis also showed that New Yorkers made best use of ranking candidates in order of preference, as they ultimately did vote for their favorite candidate, as shown in Table 6, where a large majority (91%) of voters ranked Mamdani. Furthermore, another large majority (82%) of voters said that they ranked at least two candidates, with 45% saying they ranked the maximum number permitted, five (SurveyUSA, 2025).

Table 6

Ranking usage by first choice

Candidate	% of voters who ranked multiple choices
Any	78%
Zohran Mamdani	91%
Andrew Cuomo	55%
Brad Lander	96%
Adrienne Adams	84%
Scott Stringer	86%
Zellnor Myrie	93%
Whitney Tilson	93%

Michael Blake	85%
Jessica Ramos	71%

Source: SurveyUSA (2025). Results of SurveyUSA Election Poll #27619. Retrieved from https://www.surveyusa.com/client/PollReport.aspx?g=edce291d-d3e8-4bb9-94f2-1c6b5e94975e

Reasons for ranking: Voters denoted that there were varied reasons why they chose to rank. These reasons stemmed from "Ranking allowed me to vote for candidates who aligned with my values." Ranking allows me to support multiple candidates/ have a say in who gets elected and more as shown in table 7.

Table 7

Reasons for ranking by percentage

Ranking allowed me to vote for candidates who aligned with my values.	58%
Ranking allowed me to support multiple candidates.	52%
Ranking gave me more of a say in who gets elected	42%
Ranking gives my vote more power	36%
Ranking makes me feel like my voice will be better heard by elected officials	24%
Ranking made me feel better about the outcome, regardless of who wins.	19%
The directions suggest I should rank candidates.	10%

Source: SurveyUSA (2025). Results of SurveyUSA Election Poll #27619. Retrieved from https://www.surveyusa.com/client/PollReport.aspx?g=edce291d-d3e8-4bb9-94f2-1c6b5e94975e

New themes in RCV outcomes, 2025

Voters who had not voted before: Meanwhile, several new themes emerged during the 2025 election cycle. For one, there was a major increase in new voters in this cycle with a large majority (60 %) of newly registered voters turning out to vote, which was double the number of voters who had been registered for more than a year (Hernandez, 2025). The second theme was that younger voters ages 18 to 29 had the highest turnout of any age group at 35.2%, about twice their turnout in 2021 at 17.9%. This was compared with voters aged 80 and older, who had the lowest turnout at 20.9 percent, a rate similar to past years. And thirdly, turnout looked different across the five boroughs. During this election, Manhattan led the city with 45% percent turnout, up seven points from 2021, while Staten Island had the lowest turnout at 16% (Hernandez, 2025).

Another new theme that emerged this election cycle was in the Variety of Voting Methods used in the Primary: During this 2025 cycle, voters used multiple voting methods, such as voting in person on election day, voting early, by mail, or by absentee ballot. While the largest majority (56.9%) voted in person on Election Day, 34.5 percent voted early in person, while other minor percentages (3.6%) voted early by mail, (3.4%) voted by absentee ballot, and 1.7% used other methods, such as affidavit or military ballots (Hernandez, 2025).

Campaign strategies: Meanwhile, candidates continued to endorse and support each other, as RCV meant that ranking a candidate second or even third could mean that the candidate could win in RCV elections. And as anticipated, Zohran Mamdani and Brad Lander, the highest-ranking progressive candidates in the race, endorsed each other one day after the final primary debate, during which both attacked Cuomo (Hutchinson, 2025). In backing each other, the two candidates each urged their supporters to rank the other second as they seek every advantage before early voting began. Mamdani famously said, “Don’t just rank me — fill out the rest of your ballot and rank Lander, Adams, and Myrie instead. He delivered the guidance in Hindi and used glasses of mango lassi to demonstrate how votes are distributed in the city’s ranked method (Politico, 2025). At the end, the two prominent candidates, Zohran Mamdini and Andrew Cuomo, faced each other, and after eleven rounds of counting, Zohran Mamdini won the race, becoming the city's first Muslim mayor elected with a majority consensus, as shown in Table 8.

Table 8

New Yor City Democratic mayoral primary results

Candidate	Round 1	Round 10	Round 11
Zohran Mamdini	43.82%	46.41%	56.39%
Andrew Coumo	36.12%	39.17%	43.61%
Brad Lander	11.26%	14.41%	--
Adreinne Adams	4.12%	--	--
Scott stinger	1.66%	--	--
Zellnor Myrie	.99%	--	--
Whitney Tilson	.79%	--	--
Micheal Blake	.41%	--	--
Jessica Ramos	.4%	--	--
Paperboy Price	.15%	--	--
Selma Bartholomew	.14%	--	--
Write ins	.15%	--	--

Source: Hutchinson, R & Mohamud, A (2025). Retrieved from https://fairvote.org/report/nyc-report-2025/?section=executive-summary

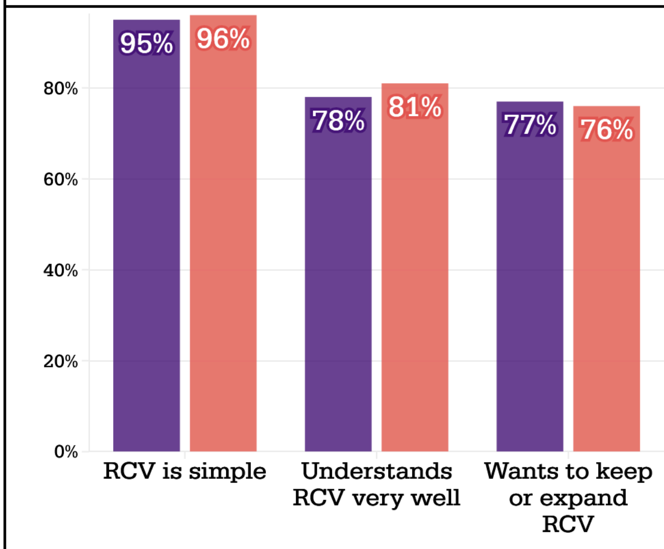
Comparing RCV use in New York: 2021 to 2025

RCV processes: The future of RCV in New York is bright from New Yorkers perspective as 1.1 million people cast a ballot making RCV their new normal. Most especially as the vast majority of New Yorkers (76% percent) in 2025 would like to keep or expand ranked choice voting in the city's elections, as they understood and appreciated the ranking method in comparison to the 77% affirming the same in 2021, as shown in Figure 2.

Understanding RCV processes: Other affirmations of the method include 81% of voters denoting their understanding or extremely well understanding of the RCV process in comparison to 78% percent affirming the same in 2021, as shown in the same figure 2. Relatedly, these figures remain consistent over the years, with 96% percent in 2025 finding RCV simple to use, versus the 95% percent finding the same, as shown in the same Figure 2.

Figure 2:

Understanding RCV processes. New York City polling 2021 v. 2025



Source: New York City Board of Election, Citizens Union Retrieved from: https://fairvote.org/report/nyc-report-2025/?section=executive-summary

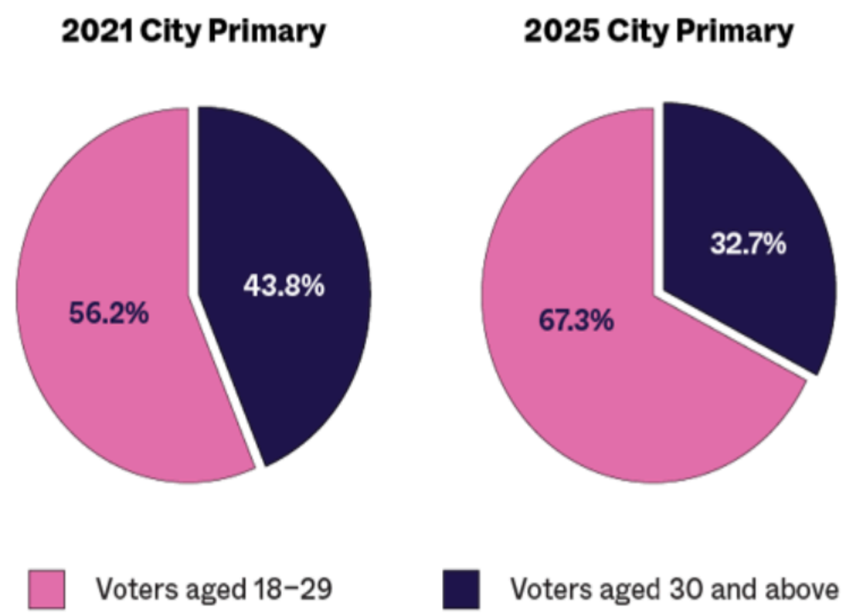
Turnout: Similarly, turnout across each election cycle with RCV soared. Take, for instance, in New York City's 2021 primaries, the city had its highest turnout in over thirty years (N.Y. City Campaign Finance Board, 2022). While during the New York 2025 mayoral election, voters were galvanized with voter turnout being the highest in half a century, as more than two million New York voters cast ballots, the most in a mayoral race since 1969, as shown in Figure 3. This surge in votes was even more notable, as the electorate had shrunk from 5.6 million registered voters four years ago to 5.3 million (Hernandez, 2025).

Registration and Turnout of younger adults: Historically, registration and turnout of younger adults have always been low throughout the country. Yet RCV changed that trend as voters in New York aged 18 to 29 made up 67.3 of new registrations, a big jump from 56.2 percent in 2021 (Hernandez, 2025). Moreover, not only did young adults register to vote, but they turned out to vote in large numbers as well. As shown in Figure 3, 56.2% of voters aged 18-29 turned out to vote in 2021, versus 67.56% in 2025, versus 67.3% in

2021. Similarly, 43.8% of voters aged 30 and above voted in 2021, versus 32.7% in 2025. In fact, turnout in 2025 was at its highest as well with 25% of first-time voters using RCV (Hernandez, 2025).

Figure 3

Voter Turnout in NYC's 2025 Primary Hits Decade-High



Source: Hernandez, M (2025). Retrieved from https://www.nycvotes.org/news-and-press/articles/2025-primary-election-turnout-soars/

Diversity Another theme that emerged was that across each iteration of RCV in New York a diverse city council was elected. Take for instance in 2021, New York elected its first-ever majority-women City Council and second-ever Black mayor following its first RCV primaries. Pursuant to 2021, they maintained a majority-women City Council in 2023. The trend held true again in 2025 with the election of a diverse city council and their first Muslim mayor (FairVote, 2025).

In fact, New Yorkers were overwhelmingly enthusiastic about ranking their votes for the first time. The Executive Director of the Center for Law and Social Justice at Medgar Evers College noted.

> "Unlike the single-winner system, RCV allows for multiple women candidates or candidates of color without concerns of vote splitting. This inclusive approach helps avoid pigeonholing candidates based on gender or race and encourages elected officials to address the needs and concerns of Black and Brown communities for re-election. RCV was a win for New York City in 2021, and we're confident that the system will continue strengthening fair representation, community engagement, and inclusive democracy during the upcoming election cycle." (Common cause, 2023).

Conclusion

Ranked-choice voting is an increasingly popular way for cities to run their elections, having grown from just 10 cities in 2016 to 50 cities, counties, and states today. Notably, five states in Nevada, Alaska, Wyoming, Kansas, and Hawaii use RCV for their presidential primaries. Furthermore, data from across the country has shown several notable outcomes which range from distinctive campaign strategies, positive campaigns, eliminated or lower vote splitting, and lower election administrative costs.

Distinctive Campaign strategies: One unique characteristic in RCV elections is that candidates can benefit from being the backup choice of voters who rank another candidate first. The reason is logical: cross-endorsements help like-minded candidates stay in the race without fear of "splitting the vote." This is compared to our current electoral system, in which ideological allies often hurt each other by splitting potential supporters.

Yet with RCV cross-endorsements benefit candidates, and voters alike, as candidates can stay in the race and make their unique case to the electorate, and voters have greater choice. As a result, several pairs of candidates have "cross-endorsed" each other in RCV contests; these candidate pairs encouraged their supporters to rank them first and second. Moreover, cross-endorsements and/or joint campaign activities RCV has shown to promote campaign civility, prevent vote-splitting, and lead to better representation. As the data show, these tools boost candidates' performance, even if they don't always translate into wins (Hutchinson & Fitzgerald, 2025).

For example, in Portland's 2024 District 3 City Council race, candidates Angelita Morillo and Tiffany Koyama Lane cross-endorsed one another and actively campaigned together. Both won seats in their three-member district, and both were the most popular second choice among voters who ranked the other first. Likewise, during Alaska's 2022 special congressional election Sarah Palin changed her strategy and cross-endorsed fellow Republican Nick Begich. When Begich was eliminated, 67% of his supporters' ballots transferred to Palin, compared to 50% in the special election. This improved Palin's performance, although she eventually lost.

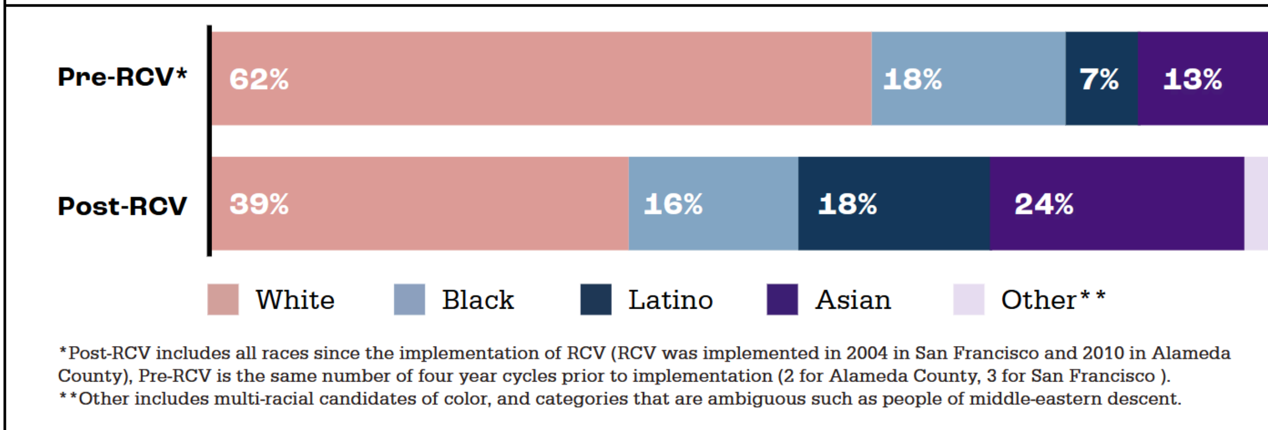
Similarly, in New York's 2021 RCV election, mayoral candidates Andrew Yang and Kathryn Garcia jointly campaigned, with Yang endorsing Garcia as a second choice. Yang's endorsement proved to be fruitful, as when he was eliminated, 45% of his ballots transferred to Garcia, thus helping Garcia win 33% more ballots than she otherwise would have. These votes saved Garcia from elimination in the seventh round, where she edged Maya Wiley by one point four percentage points. In the next round, Garcia lost to Eric Adams by less than one point (Hutchinson & Fitzgerald, 2025). Likewise, it was the case in New York's 2023 election in District 9, there was only partial crossover voting between Salaam and Taylor. As Salaam commented, about Taylor and himself, we are "together in so many things" (Otis, 2024). While in 2025, Zohran Mamdani and Brad Lander endorsed each other one day after the final primary debate, during which both attacked Cuomo.

Incentivizes positive and issue-focused campaigns. Another unique characteristic of RCV is that, as candidates compete for second-choice votes from their opponents' supporters, it disincentivizes candidates from running negative campaigns. In contrast to our current system, where candidates often attack their opponents in an attempt to highlight their differences, yet in RCV contests, candidates do best when they reach out to appeal to as many voters as possible, including those supporting their opponents, in hopes of earning a second or third spot on a voter's ballot.

Benefits women and those of color: Data from RCV elections reveal that more candidates of color were elected to office under RCV than the cities' former election systems. For instance, a study of four racially diverse Bay Area cities: San Francisco, Oakland, San Leandro, and Berkeley found a strong correlation between the racial demographics of electorates and the election of people of color by those electorates. Overall, people of color and women won far more seats since their cities started using RCV (FairVote, 2019). As shown in Table 9, people of color went from winning only 38% of seats pre-RCV elections to fifty-eight percent post-RCV (FairVote, 2019).

Table 9

Ranked choice voting and racial minority voting rights.



Source: Fair vote (2019). Retrieved from https://fairvote.app.box.com/s/npiujexebhl8ari7c61v90af3wwwfqvq

Eliminates vote splitting and spoilers. And yet another unique characteristic in RCV elections is that multiple candidates can run without having to worry about spoiling the election. In our current system, the "spoiler effect" has long been a point of contention in close political contests, where a third candidate appears to have drawn first-choice votes away from one candidate in a contested race. Yet with RCV multiple candidates are encouraged to run doing away with the spoiler effect.

Reduces the cost of running an election. RCV elections reduce costs by eliminating the need for runoffs and costly two-tiered campaigns, saving both time and money. Thus, with RCV, voters enjoy the benefits of two or more rounds of voting in a single election. And this is why single-winner RCV is also known as “instant runoff voting.”

On balance, RCV has been shown to have several appealing electoral reform qualities. Among the most notable benefits of RCV are that RVC promotes more civil campaigns and greater representation. Research suggests elections with RCV increase the odds of candidates of color and women candidates being elected. Additionally, there is compelling evidence that most citizens like using RCV as they find it easy and straightforward to use.

2026 and RCV

This year (2026) RCV will be used in nineteen cities and states across the nation, yet already there have been several notable primaries. For instance, the state of Maine, who have successfully used RCV since 2018, have already made headlines for their use of RCV for their open congressional race in ME-2, a swing district and for a hotly contested Democratic primary for the right to challenge Senator Susan Collins. And later this year Maine will use RCV again for their June 2026 presidential primaries similar to their past use of RCV for presidential primaries. Relatedly, the nation’s capital, Washington, D.C., will use RCV for the first time in its June 2026 primary in their deep-blue District, where eight Democrats are running to replace retiring Mayor Muriel Bowser.

With RCV being used by approximately seventeen million Americans across the nation the future of RCV is optimistic. Aside from RCV being used in cities and states across the nation, the Academy Awards have used RCV since 2009, for them for Best Picture award when the category was expanded to ten films. Since then, Best Picture voters have had a wider range of choices and the ability to rank their favorite films of the year in order of preference.

Certainly, as more cities, states and other organizations utilize RCV more data will be available to assess the benefits and or drawbacks of RCV. In addition, several dynamics will be worth watching in future elections such as if voters continue to consistently rank multiple candidates in each election cycle. How will candidates continue to coalesce in RCV elections? Will turnout continue to grow, especially among young voters who have had the lowest turnout over the years? Will Congress pass House Bill 6589 (https://www.congress.gov/bill/119th-congress/house-bill/6589/all-info) or Senate Bill S.3425 (https://www.congress.gov/bill/119th-congress/senate-bill/3425/all-info) this year for primary and general elections for the U.S. House and Senate?

References

Chadha, Anita. "Rank Choice Vote: An idea whose time has come." *National Social Science* (2019): 36.

Board of elections in the City of New York (2023). Primary election officials ranked choice rounds. Retrieved from https://www.vote.nyc/sites/default/files/pdf/election_results/2023/20230627Primary%20Election/rcv/025756_1.html

Common cause (2023). New Yorkers: Don't Forget to Rank in the Upcoming City Council Elections! Retrieved from https://www.commoncause.org/new-york/press/new-yorkers-dont-forget-to-rank-in-the-upcoming-city-council-elections/

Common cause (2025). La votación por orden de preferencia unió a los neoyorquinos en las primarias demócratas de este año. Retrieved from https://www.commoncause.org/es/articulos-2/ranked-choice-voting-brought-new-yorkers-together-in-this-years-democratic-primary/

FairVote (2019). Ranked choice voting and racial minority voting rights. Retrieved from https://fairvote.app.box.com/s/npiujexebhl8ari7c61v90af3wwwfqvq

FairVote (2021). New York City's Ranked Choice Voting Rollout: Better Elections Yield Better Results. Retrieved from https://fairvote.org/new_york_city_s_ranked_choice_voting_rollout_better_elections_yield_better_results/

FairVote (2023). What we learned from New York city's second ranked choice voting election. Retrieved from https://fairvote.app.box.com/s/64k7196cqg7vxz0fm7duk72b2y351225

Houghton, A. (2021). Eric Adams Holds Sizeable Lead in Latest Ranked Choice Voting Poll. Fair vote.

Hutchinson & Mohamud (2025). Ranked choice voting in New York City's 2025 primaries: an in-depth analysis. Retrieved from https://fairvote.org/report/nyc-report-2025/

Hernandez, M. (2025). Voter Turnout in NYC's 2025 Primary Hits Decade-High. New York city votes. Retrieved from https://www.nycvotes.org/news-and-press/articles/2025-primary-election-turnout-soars/

Hutchinson, R & Fitzgerald, B. (2025). Impact of Cross-Endorsements in Ranked Choice Voting Elections. Retrieved from https://fairvote.org/report/rcv-cross-endorsements-report/?section=executive-summary

Hutchinson, R. (2025). Ranked choice voting in New York City's 2025 primaries: An in-depth analysis. Retrieved from https://fairvote.app.box.com/s/2z4r6smvgse5n9gqz1f2plvsyzlmluq3

Hutchinson, R & Mohamud, A (2025). Ranked choice voting in New York City's 2025 primaries: an in-depth analysis. Retrieved from https://fairvote.org/report/nyc-report-2025/?section=executive-summary

Lamendola, Scaglia & Chan, P (2022). Why Women Won in 2021: How a Twin-Track Approach Advanced Women's Representation on the New York City Council. Retrieved from https://representwomen.app.box.com/s/p00l8z5vwfo8vru16ho0jqccsw933ay6

N.Y. votes (2025). Voter Turnout in NYC's 2025 Primary Hits Decade-High. Retrieved from https://www.nycvotes.org/news-and-press/articles/2025-primary-election-turnout-soars/

Otis, D. (2024). What we learned from New York City's second ranked choice voting election. Retrieved from https://fairvote.org/report/rcv-in-nyc-report-2023/?section=a-quieter-election-cycle

Politico (2025). Zohran Mamdani and Brad Lander are cross-endorsing in race for New York City mayor. Retrieved from https://www.politico.com/news/2025/06/13/mamdani-lander-cross-endorsing-mayor-00405357

Rank the vote NY (2021). Voter Resources. Retrieved from https://rankthevotenyc.org/resource-audience/voter-resources/

Represent women (2022). Why women won in 2021: how a twin tracked approach advanced women's representation on the New York city's council. Retrieved from https://representwomen.app.box.com/s/p00l8z5vwfo8vru16ho0jqccsw933ay6

Solomon, M. (2021). Rank the vote NYC releases Edison research exit poll on the election. Rank the vote NYC. Retrieved from http://readme.readmedia.com/RANK-THE-VOTE-NYC-RELEASES-EDISON-RESEARCH-EXIT-POLL-ON-THE-ELECTION/17989282?utm_source=newswire&utm_medium=email&utm_campaign=med ia_pr_emails

SurveyUSA (2025). Results of SurveyUSA Election Poll #27619. Retrieved from https://www.surveyusa.com/client/PollReport.aspx?g=edce291d-d3e8-4bb9-94f2-1c6b5e94975e

Codes of Ethics and Codes of Conduct Across the Social Sciences: Aspirational Norms or Regulatory Rules?

Bruce K. Bayley

Department of Criminal Justice, Weber State University

Abstract

This article presents a mixed-methods comparative analysis of 195 professional codes drawn from 13 disciplinary and governmental categories across the social sciences and public sectors in the United States. The dataset spans three primary document types: (a) Codes of Ethics (n = 102), (b) Codes of Conduct (n = 48), and (c) hybrid documents (n = 45), and covers 11 professional association disciplines and two government sectors. Descriptive statistics, comparative content analysis, and hierarchical cluster analysis were employed to address four research questions concerning dataset demographics, thematic content by document type, structural comparison across types, and cross-category thematic variation. The defining empirical finding is that Anti-Harassment is present in 62.5% of Codes of Conduct, compared to 5.9% of Codes of Ethics ($\chi^2 = 64.24$, $p < .001$), identifying Codes of Conduct as a primary institutional vehicle for #MeToo-era normative updating. Cluster analysis further reveals that government and professional association documents occupy fundamentally different normative terrain: government codes prioritize conflicts of interest (76.7%) and impartiality obligations, while professional association codes focus on equity, competence, and practitioner development. These findings carry direct implications for professional ethics education, institutional governance, and interdisciplinary standard-setting.

Keywords: professional codes; codes of ethics; codes of conduct; sociology of professions; comparative content analysis

Introduction

Background and Rationale

Professional codes serve as foundational normative frameworks that define the obligations, aspirations, and behavioral boundaries of practitioners across disciplines. Despite their prevalence in academic and government settings, systematic comparative analysis of how these documents are structured, what themes they prioritize, and how they vary across professions remains limited in the literature. The dataset analyzed here, 195 professionally ratified code documents spanning 13 categories, 11 professional association disciplines, and 2 government sectors (federal and state), collected between 2025 and 2026, offers an opportunity to address that gap directly.

The dataset includes three primary document types: Codes of Ethics (aspirational norms), Codes of Conduct (behavioral rules and enforceable standards), and hybrid or nonstandard formats. This threefold typology, recognized in professional ethics scholarship (Cava et al., 2011; Frankel, 1989), creates a natural comparative framework. Prior literature has rarely examined all three types simultaneously across a wide range of social science fields and government agencies; this article makes that contribution.

For practitioners and professional leaders, understanding how code documents are structured, and how they vary across disciplines, has direct implications for curriculum design, institutional governance, and the revision of professional standards. The findings reported here offer cross-disciplinary evidence that professional associations, educators, and policymakers can draw on when updating or benchmarking their own normative frameworks.

Significance of the Study

Understanding how professional code documents are structured and what values they prioritize is consequential for multiple audiences. Practitioners encounter these documents as defining features of professional socialization, they shape daily decision-making and occupational identity in ways that curriculum alone does not. Educators rely on them to anchor applied ethics instruction and institutional leaders consult them when governance frameworks require revision or modernization. For scholars of professional sociology and organizational ethics, codes serve as evidence of the ongoing social construction of professional boundaries (Abbott, 1988; Kjonstad & Willmott, 1995).

Research Questions

This study was guided by four interrelated research questions that moved from typological description to thematic comparison to cross-professional variation.

Research Question 1 — RQ1 (Dataset Demographics): What are the typological and demographic characteristics of professional code documents across 13 disciplinary and governmental categories, including document type distribution, geographic scope, adoption decade, and revision recency?

Research Question 2 — RQ2 (Thematic Content by Type): What are the primary themes in Codes of Ethics, Codes of Conduct, and hybrid documents, and how do those themes reflect each document type's normative orientation?

Research Question 3 — RQ3 (Structural Comparison): How are Codes of Ethics, Codes of Conduct, and hybrid documents structurally similar and different with respect to length and scope?

Research Question 4 — RQ4 (Cross-Category Thematic Variation): How do primary ethical themes vary across the 13 profession categories, and what patterns of convergence and divergence emerge along the axes of professional function and sector?

Review of Relevant Literature

The Sociology, Function, and Comparative Study of Professional Codes

The academic study of professional ethics codes dates at least to Greenwood's (1957) foundational essay on the attributes of professions, which identified a code of ethics as a defining feature distinguishing professions from occupations. Frankel's (1989) taxonomy, distinguishing aspirational, educational, and regulatory codes, remains the most widely cited typological framework in the field and directly informs this study's threefold analysis. More recently, Cava et al. (2011) examined how codes function as symbolic boundary-markers in addition to behavioral guides. Kaptein (2004) developed a virtue-based analytical framework for evaluating codes that has been extended to subsequent comparative work (Kaptein, 2011), and Abbott's (1988) sociology of professions framework posits that professions compete for jurisdictional control through, among other mechanisms, the elaboration of ethical systems.

Research in organizational sociology has examined why organizations adopt ethics codes, finding that adoption is driven by a combination of normative isomorphism — the process by which organizations adopt similar practices because of shared professional norms rather than legal mandates (DiMaggio & Powell, 1983), legitimacy-seeking (Suchman, 1995), and genuine ethical commitment. Helin and Sandström (2010) examined how a corporate ethics code was received and resisted within a multinational subsidiary, finding that empirical studies of code content and impact remain methodologically heterogeneous and under-theorized; a gap this study partially addresses through systematic comparative analysis. Singh (2011) compared ethics codes across professional associations and found substantial variation in thematic emphasis, particularly around stakeholder orientation and enforcement mechanisms.

Government Ethics, Diversity, Equity, and Inclusion (DEI), and Cross-Professional Themes

Government ethics codes occupy a distinct normative space, being legally binding rather than aspirationally framed. The literature distinguishes between ethics as compliance (rule-following) and ethics as values (character and judgment), with government codes historically emphasizing the former (Cooper, 2011). West and Berman (2004) identified significant variation in scope, enforcement, and civic culture across state government ethics programs. Menzel (2012) emphasized the distinction between codes that regulate conflicts of interest (transactional ethics) and codes that articulate public service values (aspirational ethics); a distinction that maps onto the Ethics/Conduct typology used here.

In terms of social action, previous studies have documented the role of social movements in reshaping professional ethics codes. McDonald (2009) traced how #MeToo-era disclosures accelerated anti-harassment language adoption in professional codes, and Dougherty and Smythe (2004) found that codes adopted in response to external social pressure tend to be more specific in their prohibitions, but not always more effective in practice. On diversity, equity, and inclusion provisions, Kalev et al. (2006) found that formal organizational structures for equity are associated with measurable demographic outcomes.

The literature identifies a set of near-universal themes in professional ethics codes: honesty and integrity, client/public protection, professional competence, conflicts of interest, confidentiality, and respect for persons (Frankel, 1989). In addition, Schwartz (2002) identified six core values that are present across business ethics codes. Those six core values are: (a) trustworthiness, (b) respect, (c) responsibility, (d) fairness, (e) caring, and (f) citizenship. Hugman (2013) raised postcolonial cautions worth taking seriously: dominant ethics frameworks risk marginalizing non-Western epistemologies, which means treating apparent universality as culturally neutral may itself be an interpretive error.

Methods

Dataset and Sample

The data source used in this study was designated the Professional Codes Dataset: Social Sciences USA (N = 195), a purpose-built dataset of professional code documents collected between 2025 and 2026, spanning 13 categories: 11 professional association disciplines and two government sectors (federal and state). Categories were populated using purposive sampling, prioritizing the primary professional association or government body in each domain, supplemented by significant specialty association codes and conference conduct policies (Patton, 2014). All 195 documents are publicly accessible via direct URL.

Key Variables

Doc Type Simplified is a three-category typological variable: Code of Ethics (aspirational norms; n = 102), Code of Conduct (behavioral rules and enforceable standards; n = 48), and Other Documents (hybrid documents, statements of principles, and guidelines; n = 45). Notable Revision Flag is a binary indicator of whether a document was explicitly revised in response to social justice events, #MeToo, COVID-19, or digital/AI ethics concerns; 56 documents (28.7%) carry this flag. Document type composition and geographic scope by category are presented in Tables 1 and 2.

Analytical Approach

Analysis proceeded in three phases: descriptive statistics addressing RQ1; comparative content analysis addressing RQ2–RQ4; and interpretive synthesis integrating statistical findings with the sociology of professions literature. Chi-square tests compared categorical distributions. Kruskal-Wallis tests with Bonferroni-corrected pairwise Mann-Whitney U follow-ups compared word counts given confirmed nonnormality. Ward linkage hierarchical cluster analysis was applied to the 13×19 theme prevalence matrix to identify categories with similar thematic profiles.

Thematic Coding

Thematic coding was conducted in two passes. In the first (deductive) pass, 12 codes were applied based on the work of Frankel (1989) and Schwartz (2002): Public/Client Protection, Professional Competence, Honesty/Integrity, Conflicts of Interest, Confidentiality/Privacy, Informed Consent, Non-Discrimination/Equity, Anti-Harassment, Environmental/Heritage Stewardship, Accountability/Enforcement, Digital/AI Ethics, and Community/Public Education. In the second (inductive) pass, seven emergent codes were added through systematic re-reading: Research Integrity, Social Justice/Racial Equity, Fiduciary/Financial Stewardship, Mentoring/Supervisory Ethics, Professional Relations/Collegiality, Multicultural/Cultural Competence, and Public Trust/Service Ethic. Inter-rater reliability was established on a randomly selected subsample of 40 documents (20.5%), yielding Cohen's κ values ranging from .74 to .93 across the 12 deductive codes ($M = .82$; Landis & Koch, 1977).

Results

Dataset Demographics and Typological Overview (RQ1)

The 195-document dataset is balanced across all 13 categories (15 codes per category). Codes of Ethics constitute the largest group ($n = 102$, 52.3%), followed by Codes of Conduct ($n = 48$, 24.6%) and Other Documents ($n = 45$, 23.1%). Document type composition varies widely by category (see Table 1): U.S. Government Agencies was the sole category dominated by Codes of Conduct (9 of 15, 60.0%), while Anthropology and Archaeology and Child and Family Studies each contain 12 Codes of Ethics. Geographic scope, shown in Table 2, was predominantly national (58.5%), with 29.2% international and 12.3% regional or state; scope did not differ significantly across document types ($\chi^2(4) = 8.41$, $p = .078$), indicating that scope was driven by disciplinary category rather than document format.

Table 1

Document Type Distribution by Category (N = 195)

Category	Code of Ethics	Code of Conduct	Other	Total	% CoE

Anthropology & Archaeology	12	3	0	15	80.0
Child & Family Studies	12	2	1	15	80.0
Communication Studies & Media	11	2	2	15	73.3
Criminology & Criminal Justice	10	2	3	15	66.7
Economics & Behavioral Economics	2	7	6	15	13.3
Education	7	3	5	15	46.7
Philosophy	3	3	9	15	20.0
Political Science & Public Admin.	7	4	4	15	46.7
Psychology	10	3	2	15	66.7
Social Work & Human Services	8	0	7	15	53.3
Sociology & Demography	8	5	2	15	53.3
State Government Agencies	8	5	2	15	53.3
U.S. Government Agencies	4	9	2	15	26.7
Total	**102**	**48**	**45**	**195**	**52.3**

Note. CoE = Code of Ethics. % CoE = percentage of 15 category documents classified as Code of Ethics.

Table 2

Geographic Scope Distribution by Category (N = 195)

Category	International (n)	National (n)	Regional/State (n)	% International
Anthropology & Archaeology	9	5	1	60.0
Child & Family Studies	5	10	0	33.3
Communication Studies & Media	3	12	0	20.0
Criminology & Criminal Justice	5	10	0	33.3
Economics & Behavioral Economics	7	7	1	46.7
Education	1	14	0	6.7
Philosophy	9	6	0	60.0
Political Science & Public Admin.	10	4	1	66.7
Psychology	3	12	0	20.0

Social Work & Human Services	1	14	0	6.7
Sociology & Demography	4	5	6	26.7
State Government Agencies	0	0	15	0.0
U.S. Government Agencies	0	15	0	0.0
Total / Dataset Mean	**57 (29.2%)**	**114 (58.5%)**	**24 (12.3%)**	**29.2**

Note. Chi-square test of scope by document type: $\chi^2(4) = 8.41$, $p = .078$ (not significant). Scope was driven primarily by disciplinary category, rather than document format.

Post-2020 revision rates varied widely by category, with State Government Agencies (86.7%) and Child and Family Studies (60.0%) being the most predominant, while Criminology and Criminal Justice showed the highest proportion of unrevised documents (66.7%). The 2010s (n = 47, 24.1%) and 1990s (n = 41, 21.0%) were the two peak adoption decades, together accounting for 45.1% of the dataset.

Primary Themes by Document Type (RQ2)

Table 3 presents the theme-frequency matrix for all 19 codes across document types. Across the full dataset, the three most prevalent themes were Accountability/Enforcement (37.9%), Non-Discrimination/Equity (37.4%), and Honesty/Integrity (34.4%), consistent with research identifying these as near-universal contemporary professional values (Kaptein, 2011; Schwartz, 2002). These three themes function as a normative constant shared across all document types.

Table 3

Selected Theme Frequencies by Document Type: Deductive (D) and Emergent (E) Codes

Code	**Theme (Type)**	**CoE % (n=102)**	**CoC % (n=48)**	**Other % (n=45)**	**All % (n=195)**	**Sig.**
PC	Public/Client Protection (D)	29.4	12.5	2.2	19.0	***
PQ	Professional Competence (D)	31.4	6.2	15.6	21.5	**
HI	Honesty/Integrity (D)	42.2	20.8	31.1	34.4	*
CI	Conflicts of Interest (D)	16.7	18.8	8.9	15.4	ns
NDE	Non-Discrimination/Equity (D)	34.3	41.7	40.0	37.4	ns
AH	Anti-Harassment (D)	5.9	62.5	13.3	21.5	***
AE	Accountability/Enforcement (D)	36.3	54.2	24.4	37.9	*
DA	Digital/AI Ethics (D)	19.6	22.9	20.0	20.5	ns
CPE	Community/Public Education (D)	21.6	6.2	6.7	14.4	*
RI	Research Integrity (E)	13.7	6.2	17.8	12.8	ns
SJ	Social Justice/Racial Equity (E)	9.8	8.3	22.2	12.3	ns

ME	Mentoring/Supervisory Ethics (E)	14.7	6.2	20.0	13.8	ns
PT	Public Trust/Service Ethic (E)	11.8	8.3	6.7	9.7	ns

Note. CoE = Code of Ethics; CoC = Code of Conduct. D = Deductive code; E = Emergent code. Sig. = significance of 2 × 3 chi-square test: *** $p < .001$; ** $p < .01$; * $p < .05$; ns = not significant. Anti-Harassment and Public/Client Protection survive Bonferroni correction (corrected $\alpha = .0026$). Six codes (CP, IC, EHS, FID, PR, MCF) omitted for brevity; full matrix available from the author.

Codes of Ethics were characterized by a practitioner-centered thematic profile: the five leading themes were Honesty/Integrity (42.2%), Accountability/Enforcement (36.3%), Non-Discrimination/Equity (34.3%), Professional Competence (31.4%), and Public/Client Protection (29.4%). Professional Competence was significantly more common in Codes of Ethics than in Codes of Conduct (31.4% vs. 6.2%; $\chi^2(2) = 13.43$, $p = .001$), and Informed Consent was entirely absent from Codes of Conduct, while present in 13.7% of Codes of Ethics.

The defining finding of this study is that Anti-Harassment was present in 62.5% of Codes of Conduct, compared to 5.9% of Codes of Ethics and 13.3% of Other Documents ($\chi^2(2) = 64.24$, $p < .001$). This finding constitutes the strongest empirical marker distinguishing Codes of Conduct from other document types, reflecting their role as a primary institutional vehicle through which #MeToo-era anti-harassment norms have been codified across professions (Dougherty & Smythe, 2004; McDonald, 2009). Accountability/Enforcement was also significantly higher in Codes of Conduct (54.2% vs. 36.3%; $p = .011$). Other Documents show elevated Social Justice/Racial Equity (22.2%) relative to both Codes of Ethics (9.8%) and Codes of Conduct (8.3%), suggesting that activist-oriented organizations articulate social justice commitments through nontraditional document forms (Singh, 2011).

Structural Comparison Across Document Types (RQ3)

Table 4 synthesizes the structural comparison across document types with respect to length and scope. Median word counts follow a clear hierarchy: Code of Ethics = 2,000 words; Code of Conduct = 1,000 words; Other = 700 words, confirmed by a significant Kruskal-Wallis test ($H = 6.82$, $p = .033$). The normative orientation distinction is equally pronounced: Codes of Ethics are aspirational (emphasizing practitioner character and professional duty), while Codes of Conduct are regulatory (emphasizing prohibited behaviors and enforcement mechanisms). Other/hybrid documents occupy an intermediate position, often highlighting organizational mission alongside behavioral standards.

Table 4

Structural Comparison Across Document Types: Length and Scope

Dimension	Code of Ethics (n = 102)	Code of Conduct (n = 48)	Other/Hybrid (n = 45)
Normative orientation	Aspirational: practitioner character and professional duty	Regulatory: prohibited behaviors, enforcement mechanisms	Mixed: spans both; often foregrounds organizational mission
Universal themes (≥20% in all types)	Honesty/Integrity (42.2%); Non-Discrimination (34.3%); Accountability (36.3%)	Non-Discrimination (41.7%); Accountability (54.2%); Honesty/Integrity (20.8%)	Non-Discrimination (40.0%); Honesty/Integrity (31.1%); Accountability (24.4%)
Most distinctive theme	Public/Client Protection: 29.4% vs. CoC 12.5% (χ^2 = 16.75, $p < .001$)	Anti-Harassment: 62.5% vs. CoE 5.9% (χ^2 = 64.24, $p < .001$)	Social Justice: 22.2% vs. CoE 9.8% (trend, $p = .067$)
Median word count	2,000 words	1,000 words	700 words

Note. Word count distributions significantly different across types: Kruskal-Wallis $H(2) = 6.82$, $p = .033$.

Thematic Variation Across Profession Categories (RQ4)

Hierarchical cluster analysis of the 13 × 19 theme prevalence matrix identified four interpretable clusters of categories with similar thematic profiles (see Table 5).

Table 5

Four-Cluster Solution: Category Groupings and Distinguishing Themes

Cluster	Categories (n)	Top Themes	Interpretation
A — Government Regulatory	State Government Agencies; U.S. Government Agencies (n = 30)	Conflicts of Interest (76.7%); Honesty/Integrity (53.3%); Public Trust (26.7%)	Ethics as statutory compliance; conflict-of-interest and impartiality obligations dominate; social justice absent
B — Accountability-Oriented	Communication, Criminology, Economics, Philosophy, Political Science, Sociology (n = 90)	Accountability (46.7%); Non-Discrimination (40.0%); Anti-Harassment (30.0%)	Event-based accountability norms; #MeToo-era anti-harassment codes prominent; digital ethics highest in Communication
C — Practice & Care Professions	Child & Family, Education, Psychology, Social Work (n = 60)	Non-Discrimination (46.7%); Competence (40.0%); Mentoring Ethics (30.0%)	Client-centered duty of care; competence and confidentiality obligations; mentoring ethics elevated
D — Heritage & Field Sciences	Anthropology & Archaeology (n = 15)	Heritage Stewardship (60.0%); Community Education (40.0%); Collegial Relations (40.0%)	Unique ecological niche; Indigenous rights and repatriation define the ethical terrain

Note. Top themes reported as percentage of documents within each cluster. Four-cluster solution derived from Ward linkage hierarchical cluster analysis applied to the 13 × 19 theme prevalence matrix. Anthropology & Archaeology (Cluster D) is isolated from all other categories, reflecting its unique thematic profile.

Cluster A (Government Regulatory) groups the two government sectors, separated from all professional association categories. Their signature theme was Conflicts of Interest (76.7% government vs. 4.2% professional associations), the largest between-sector difference in the dataset. Public Trust/Service Ethic was also significantly elevated in government documents (26.7% vs. 6.7%), while Social Justice/Racial Equity was entirely absent (0.0% vs. 14.5%). These contrasts confirm that government and professional association codes occupy fundamentally different normative terrain: government codes primarily regulate conflicts of interest and impartiality obligations, while professional association codes focus on equity, justice, and practitioner development (Cooper, 2011; West & Berman, 2004).

Cluster B (Accountability-Oriented Disciplines) includes Communication Studies and Media, Criminology and Criminal Justice, Economics and Behavioral Economics, Philosophy, Political Science and Public Administration, and Sociology and Demography; disciplines sharing high Accountability/Enforcement prevalence (cluster mean = 46.7%) and elevated Anti-Harassment (30.0%). Communication Studies and Media stood out for its exceptionally high Digital/AI Ethics prevalence (60.0%), reflecting journalism's early engagement with platform ethics and AI-generated content governance.

Cluster C (Practice and Care Professions) groups Child and Family Studies, Education, Psychology, and Social Work and Human Services. Their distinguishing themes are Non-Discrimination/Equity (cluster mean = 46.7%), Professional Competence (40.0%), and Mentoring/Supervisory Ethics (30.0%). Psychology showed the highest Professional Competence prevalence of any category (60.0%), anchored in clinical licensing obligations; Education showed the highest Mentoring/Supervisory Ethics prevalence (53.3%), reflecting concern with faculty-student power dynamics.

Cluster D (Heritage and Field Sciences) contains only Anthropology and Archaeology, isolated by its Environmental/Heritage Stewardship prevalence (60.0% vs. dataset mean of 7.7%); the most profession-specific finding in the entire analysis. The helping professions cluster differs significantly from the research/knowledge disciplines cluster on seven of 19 codes; helping professions were significantly higher on Professional Competence (46.7% vs. 13.3%) and Confidentiality/Privacy (31.1% vs. 13.3%), consistent with Abbott's (1988) claim that helping professions frame ethics around obligations to individual clients, while research professions frame ethics around the integrity of inquiry and governance of disciplinary communities.

Discussion

This study's four research questions were designed to move from descriptive typology toward cross-professional comparison. What the data returned was, in some respects, more orderly than anticipated, and in others, more uneven. Professional ethics documents share a stable normative core, but beneath that surface consensus, document type and professional function produce sharp variation, and the response to external sociopolitical pressure varies considerably across fields.

The Aspirational–Regulatory Divide

The most consistent finding across RQ2 and RQ3 was that Frankel's (1989) aspirational–regulatory distinction holds empirically at the level of content, not just form. Codes of Ethics are longer and characteristically anchored in practitioner-centered themes (Professional Competence, Public/Client Protection, and Informed Consent) that together present a vision of what a good practitioner looks like. Codes of Conduct are shorter, more prohibitory, and organized around enforcement. The implication is that the two document types serve genuinely different institutional functions: Codes of Ethics socialize practitioners into professional identity, whereas Codes of Conduct operationalize behavioral compliance and respond reactively to social crises. They are not interchangeable, and treating revision of one as a substitute for the other is unlikely to achieve both goals.

Sectoral and Disciplinary Differentiation

The government/professional association divide identified in the cluster analysis replicates and extends findings from the public administration ethics literature (Cooper, 2011; West & Berman, 2004). The near-total absence of Social Justice/Racial Equity language in government documents (0.0%), alongside the overwhelming prevalence of Conflicts of Interest (76.7%), reflects the historically transactional character of public sector ethics. This contrast is not merely semantic. Government codes are designed to prevent self-dealing and protect institutional impartiality. As such, their operative question is whether a public official can be trusted to act without personal interest. Professional association codes operate from a different premise entirely: they articulate practitioner identity and advance field-specific visions of what the profession is for. The two categories are answering different questions, which is why importing language freely between them rarely works as intended.

Within the professional association sector, the cluster analysis revealed that disciplinary function, specifically the helping/research distinction, is a more powerful organizer of thematic profiles than disciplinary content area. This supports Abbott's (1988) prediction that professions encode their jurisdictional claims ethically: helping professions' ethical focus is defined by the dyadic obligation to individual clients, while knowledge professions' ethical emphasis is defined by the integrity of inquiry and the governance of disciplinary communities.

The Role of Sociopolitical Change

A consistent theme across the results was that the current wave of professional ethics revision is externally driven rather than internally generated. The concentration of Anti-Harassment in Codes of Conduct collectively suggests that the #MeToo movement, COVID-19, and digital/AI ethics pressures are the primary catalysts of recent normative change, not internal professional deliberation. Dougherty and Smythe (2004) found that externally pressured codes tend toward greater specificity in their prohibitions, a pattern supported by this study. Whether that specificity translates into broader ethical vision, or merely reactive rule-writing, is a question the current data cannot resolve. The high proportion of unrevised documents in Criminology and Criminal Justice (66.7%) stands out given the field's central concern with justice and accountability, and warrants attention from professional leadership.

Limitations and Directions for Future Research

Two limitations constrain interpretation here. The more significant, in my view, is the binary coding approach: treating all detected themes as equivalent in prominence regardless of how much of a document is devoted to them flattens real differences in emphasis. A code that devotes a single clause to honesty and twelve pages to conflict-of-interest procedures is coded identically to one that centers honesty throughout. That bluntness is an artifact of the methodology that future work should address through weighted or proportional coding. The second limitation is geographic: restricting the dataset to English-language, U.S.-based documents precludes generalization to non-English professional ethics traditions. Hugman (2013) cautions that treating near-universal themes as culturally neutral risks marginalizing non-Western points of view, a finding that applies directly to a dataset limited by a single national and linguistic context.

Both limitations point toward manageable solutions. The geographic and linguistic constraints are the more straightforward to address: expanding the dataset to include non-English professional ethics traditions would allow direct testing of whether the thematic clusters identified here are culturally specific to the U.S. context or approximate something more universal. The methodological limitation is harder to fix, but arguably more important: the field would benefit from weighted or proportional coding approaches that distinguish a document that devotes three paragraphs to a theme from one that gives it a single clause. Beyond the dataset itself, practitioner reception remains almost entirely unstudied; survey and interview methods could assess whether exposure to specific thematic content is actually associated with practical ethical decision-making (Lere & Gaumnitz, 2003).

Conclusion

Cross-disciplinary comparison of professional code documents is harder than it looks. The 195 documents analyzed here are not, at first glance, very different from one another: most invoke honesty, nondiscrimination, and accountability, written in the earnest register of institutional aspiration. What the analysis reveals is that beneath this surface consensus, document type and professional sector are powerful organizers of what codes actually do. Codes of Ethics socialize practitioners into professional identity; Codes of Conduct operationalize compliance and absorb the shock of social crises. Government codes regulate self-dealing; professional association codes articulate visions of practice. These are not stylistic differences. Instead, they reflect genuinely different theories of what code documents are for. These findings carry implications for how professional associations benchmark their documents against peers, how educators frame ethics instruction, and how scholars of the professions interpret normative convergence. Future research holds the promise of more nuanced distinctions: weighted content analysis, cross-national comparison, and practitioner-level reception studies would each extend what systematic comparison of the codes themselves can establish.

References

Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226189666.001.0001

Cava, A., Murray, B., & Webber, S. (2011). The professionalization of compliance. *American Business Law Journal, 48*(3), 1–54. https://doi.org/10.1111/j.1744-1714.2011.01113.x

Cooper, T. L. (2011). *The responsible administrator: An approach to ethics for the administrative role* (5th ed.). Jossey-Bass.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147–160. https://doi.org/10.2307/2095101

Dougherty, D. S., & Smythe, M. J. (2004). Sensemaking, organizational culture, and sexual harassment. *Journal of Applied Communication Research, 32*(4), 293–317. https://doi.org/10.1080/0090988042000275998

Frankel, M. S. (1989). Professional codes: Why, how, and with what impact? *Journal of Business Ethics, 8*(2–3), 109–115. https://doi.org/10.1007/BF00382575

Greenwood, E. (1957). Attributes of a profession. *Social Work, 2*(3), 45–55. https://doi.org/10.1093/sw/2.3.45

Helin, S., & Sandström, J. (2010). Resisting a corporate code of ethics and the reinforcement of management control. *Organization Studies, 31*(5), 583–604. https://doi.org/10.1177/0170840610372292

Hugman, R. (2013). *Culture, values and ethics in social work: Embracing diversity*. Routledge. https://doi.org/10.4324/9780203094907

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review, 71*(4), 589–617. https://doi.org/10.1177/000312240607100404

Kaptein, M. (2004). Business codes of multinational firms: What do they say? *Journal of Business Ethics, 50*(1), 13–31. https://doi.org/10.1023/B:BUSI.0000021051.53460.da

Kaptein, M. (2011). Toward effective codes: Testing the relationship with unethical behavior. *Journal of Business Ethics, 99*(2), 233–251. https://doi.org/10.1007/s10551-010-0652-5

Kjonstad, B., & Willmott, H. (1995). Business ethics: Restrictive or empowering? *Journal of Business Ethics, 14*(6), 445–464. https://doi.org/10.1007/BF00872099

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Lere, J. C., & Gaumnitz, B. R. (2003). The impact of codes of ethics on decision making: Some insights from information economics. *Journal of Business Ethics, 48*(4), 365–379. https://doi.org/10.1023/B:BUSI.0000005783.12302.b0

McDonald, G. (2009). An anthology of codes of ethics. *European Business Review, 21*(4), 344–372. https://doi.org/10.1108/09555340910970445

Menzel, D. C. (2012). *Ethics management for public administrators: Building organizations of integrity* (2nd ed.). Routledge. https://doi.org/10.4324/9781315704500

Patton, M. Q. (2014). *Qualitative research and evaluation methods* (4th ed.). SAGE.

Schwartz, M. S. (2002). A code of ethics for corporate code of ethics. *Journal of Business Ethics, 41*(1–2), 27–43. https://doi.org/10.1023/A:1021393904930

Singh, J. (2011). Determinants of the effectiveness of corporate codes of ethics: An empirical study. *Journal of Business Ethics, 101*(3), 385–395. https://doi.org/10.1007/s10551-010-0727-3

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571–610. https://doi.org/10.2307/258788

West, J. P., & Berman, E. M. (2004). Ethics training in U.S. cities: Content, pedagogy, and impact. *Public Integrity, 6*(3), 189–206. https://doi.org/10.1080/10999922.2004.11051235

Human Trafficking Awareness and Preparedness Among University Students: Implications for Healthcare Education and Victim Identification

Anthony P. LaRose, *University of Tampa*
Yelitza Colmenares, *University of Tampa*
Sean Maddan, *Texas A&M International*
Chris Gurrie, *University of Tampa*

Abstract

Human trafficking remains significantly underidentified within healthcare settings despite frequent contact between victims and medical professionals. Nurses and physician assistants (PAs), as frontline providers, are uniquely positioned to recognize indicators of exploitation; however, identification is often hindered by gaps in awareness, persistent misconceptions, and limited preparedness. This study examines human trafficking awareness, attitudes and misconceptions regarding victim agency, empathic responses toward trafficked individuals, and perceived confidence to intervene among university students, with particular attention to students preparing for patient-facing healthcare roles. Using survey data from 179 undergraduate and graduate students at a private university in the southeastern United States, including a 33-student nursing/physician assistant subgroup, analyses assessed associations between demographic characteristics, academic discipline, prior educational exposure, and trafficking-related perceptions. Results indicate that students demonstrated strong awareness of core trafficking definitions and high emotional engagement, but misconceptions persisted, particularly regarding victim choice and ethical intervention. Prior coursework addressing human trafficking was associated with lower self-reported lack of information and greater perceived ability to assist victims. These findings underscore the importance of structured trafficking-related education in strengthening trafficking-related awareness and perceived preparedness among future healthcare professionals. Integrating human trafficking education into healthcare-related curricula may enhance early identification, reduce missed opportunities for intervention, and promote trauma-informed responses to vulnerable populations.

Key Words: human trafficking; nursing education; physician assistant students; healthcare preparedness; victim identification

Introduction

According to the United Nations Office on Drugs and Crime (UNODC, 2024), detection of trafficking has increased by 25% since 2019, with 68,090 victims identified across 135 countries in 2022; however, these figures represent only a fraction of the estimated tens of millions affected globally (Barrick & Pfeffer, 2021). The true prevalence of trafficking is obscured by a substantial "dark figure" of cases that go unreported or unrecognized, particularly within institutional settings such as healthcare, social services, and law enforcement (Bales et al., 2015; Bales et al., 2019; Cole, 2017). Research consistently demonstrates that many trafficking victims interact with service providers without being identified, underscoring the urgent need for improved detection across professional sectors, especially healthcare (Bales et al., 2019; Seymour & Silverman, 2023).

In the United States, trends reflect a similar paradox of increased visibility paired with limited accountability. Since 2019, detected trafficking cases have risen sharply, while convictions have declined, highlighting persistent gaps in prevention, protection, and prosecution efforts (UNODC, 2024). Most identified victims are women and girls trafficked for sexual exploitation, and a majority are trafficked domestically, revealing that exploitation frequently occurs within national borders rather than through international movement (UNODC, 2024). These patterns emphasize the importance of domestic institutions, including healthcare systems, in identifying and responding to trafficking.

Healthcare professionals, particularly nurses, are uniquely positioned as frontline responders to human trafficking. Nurses' sustained patient contact, trust-building roles, and involvement in both physical and psychosocial assessments place them in a critical position to observe indicators of exploitation (Butler et al., 2018; Peck & Meadows-Oliver, 2019). An estimated 50% to 88% of trafficking victims access healthcare services during exploitation, yet identification remains rare, often due to limited awareness, misconceptions about victim profiles, and uncertainty regarding appropriate responses (Chisolm-Straker et al., 2012; Donahue et al., 2019; Rapoza, 2021). Knowledge deficits and misconceptions not only impede detection but also shape communication and relational engagement with patients, limiting opportunities for disclosure and intervention. Nursing education plays a central role in shaping the knowledge, attitudes, and communicative competencies students bring into clinical practice. Awareness of trafficking indicators, empathy toward victims, and perceived preparedness influence whether future clinicians can recognize and respond effectively to potential cases (Raker, 2020; Raker, 2023). Although much of the trafficking literature focuses on nursing students, physician assistant (PA) students share similar clinical responsibilities related to patient assessment, ethical communication, and interprofessional collaboration. As frontline healthcare providers, PA students, like nurses, will inevitably encounter trafficking victims and therefore require comparable levels of awareness and preparedness.

Accordingly, the aim of this study was to examine human trafficking awareness, attitudes, empathy, and perceived preparedness among university students, with particular attention to students preparing for patient-facing healthcare roles. Because healthcare professionals may encounter trafficking victims in clinical settings, the study compares nursing/physician assistant students with students from other academic programs and examines whether prior trafficking-related education is associated with trafficking-related awareness and perceived readiness to respond.

The study was guided by the following research questions:

RQ1: What levels of human trafficking awareness, empathy toward trafficked persons, and perceived preparedness to intervene are reported by university students?

RQ2: Do nursing/physician assistant students differ from students in other academic programs in trafficking-related awareness, empathy, misconceptions, and perceived preparedness?

RQ3: Is prior trafficking-related education associated with trafficking-related awareness, victim-centered attitudes, empathy, and perceived preparedness to respond?

Literature Review

Nurses are frequently the first point of contact for vulnerable and marginalized populations, including individuals experiencing exploitation (Peck & Meadows-Oliver, 2019). Their sustained patient interaction, involvement in both physical and psychosocial assessment, and trust-building roles place them in a unique position to observe indicators of coercion, abuse, and trafficking (Butler et al., 2018). Victims of human trafficking often do not self-identify or disclose exploitation due to fear, dependency, or trauma bonding, making provider awareness essential for detection and referral (Corthay, 2019; Quincy et al., 2020). Research consistently demonstrates that trafficking victims interact with healthcare systems while being exploited, yet identification remains rare. Studies estimate that between 50% and 88% of victims access healthcare services during exploitation, but providers frequently lack the training necessary to recognize indicators or respond appropriately (Chisolm-Straker et al., 2012; Donahue et al., 2019; Peck & Meadows-Oliver, 2019). This disconnect between opportunity and preparedness underscores the importance of formal education for frontline healthcare professionals.

Although much of the literature focuses on nurses, physician assistant (PA) students share comparable clinical responsibilities, including patient assessment, ethical communication, and interdisciplinary coordination. As future frontline providers, PA students are similarly positioned to encounter trafficking victims and therefore require comparable levels of awareness and preparedness (Quincy et al., 2020). Awareness alone is insufficient to ensure identification. Providers' attitudes, misconceptions, and levels of empathy shape how patient cues are interpreted and whether concerns are acted upon. Prior research demonstrates that misconceptions about trafficking, such as beliefs that victims can easily leave or that trafficking primarily involves physical abduction, limit the range of individuals providers are prepared to identify (Hansen & Johansson, 2025; Herrero-Villoria et al., 2022; Rapoza, 2021).
Attitudinal beliefs regarding victim agency and choice are particularly consequential. Rapoza (2021) found that nursing students and professionals were more likely to hold skeptical attitudes toward certain victim populations, reflecting persistent stereotypes about who can be trafficked. Such misconceptions reinforce narrow victim profiles and contribute to underidentification within healthcare settings (Bales et al., 2019). Empathy further influences clinical engagement. Providers who lack understanding of coercion, trauma bonding, and dependency may misinterpret victim behavior as noncompliance or choice, discouraging inquiry or intervention (Lorvinsky et al., 2023). Qualitative research with trafficking survivors highlights that healthcare providers often focus narrowly on presenting physical complaints while overlooking contextual and psychosocial indicators of exploitation, a pattern attributed to limited awareness and confidence (Lorvinsky et al., 2023).

Education has been consistently identified as a key determinant of preparedness among healthcare students and professionals. Nursing students who receive formal instruction on human trafficking demonstrate higher levels of awareness, empathy, and confidence in responding to potential cases (Donahue et al., 2019; Raker, 2020; Raker, 2023). Conversely, students without targeted education often report uncertainty regarding indicators, ethical intervention, and referral pathways (Olivieri et al., 2024).

Despite increased exposure to trafficking through media and public discourse, reliance on informal sources may reinforce misconceptions rather than correct them. Without structured education addressing coercion, victim behavior, and trauma-informed communication, students may simultaneously reject overt victim-blaming narratives while still endorsing subtler misconceptions about choice and responsibility (Hansen & Johansson, 2025; Rapoza, 2021).

Uncertainty surrounding intervention further contributes to underidentification. Although nurses and PA students may recognize trafficking as harmful, ambiguity regarding ethical boundaries and appropriate responses can discourage action, perpetuating the dark figure of unidentified victims within healthcare systems (Bales et al., 2019). These findings highlight the need for standardized educational approaches that integrate accurate knowledge, empathy, and clear guidance on trauma-informed intervention.

Collectively, prior research demonstrates that nurses and PA students are positioned to identify human trafficking victims but may lack the awareness, empathy, and preparedness necessary to do so effectively. Attitudes and misconceptions shape clinical interpretation, while educational exposure is associated with confidence and willingness to intervene. The present study builds on this literature by examining how these constructs vary across academic discipline, demographic characteristics, and prior educational exposure among university students, with particular attention to a nursing/physician assistant subgroup.

Methods

This study employed a cross-sectional survey design to examine human trafficking awareness, attitudes and misconceptions, empathy toward trafficked individuals, and perceived preparedness among university students. The study also examined whether these constructs varied across demographic characteristics, academic status, academic discipline, and prior exposure to coursework addressing human trafficking.

Study Design and Participants

The final sample consisted of 179 undergraduate and graduate students recruited from a medium-sized private university in the southeastern United States. Participants were recruited through university email distribution lists and course-based announcements that included a brief description of the study and a link to an anonymous online survey. Eligibility criteria required that participants be at least 18 years old and currently enrolled in a program of study. Participation was voluntary and anonymous.

Because the study was concerned with trafficking-related awareness and preparedness in healthcare education, nursing and physician assistant students were grouped together for analytic purposes. This subgroup included 33 students, while 146 participants were enrolled in other academic programs. The nursing/physician assistant subgroup should therefore be interpreted as a healthcare-preparation comparison group within a broader university student sample, rather than as a standalone nursing or physician assistant sample.

Table 1 presents the demographic characteristics of the sample, including academic year, major grouping, prior exposure to human trafficking coursework, race, and sex.

Table 1.

Demographic Characteristics of the Sample (N = 179)

Characteristic	n	%
Academic year		
Freshman	36	20.1
Sophomore	28	15.9
Junior	32	17.9
Senior	38	21.2
Graduate student	42	23.5
Major grouping		
Nursing / Physician assistant	33	18.4
All other majors	146	81.6
Human trafficking covered in coursework		
Yes	48	26.6
No	131	73.4
Race		
White	125	70.1
Non-White	54	29.9
Sex		
Female	131	73.2
Male	42	23.5

Note. Percentages may not total 100 due to missing responses and rounding. Nursing (n = 1) and physician assistant (n = 32) majors were combined for analytic purposes.

Instrumentation / Measures

Data were collected using a structured, self-administered survey administered electronically through Qualtrics. Prior to beginning the survey, participants reviewed an informed consent statement and indicated their consent electronically. The study received institutional review board approval through the university. Incomplete survey responses were excluded from analysis, resulting in minor variation in sample size across items.

The survey included items assessing awareness of human trafficking, attitudes and misconceptions regarding victim agency and choice, emotional and empathetic responses toward trafficked individuals, and perceived confidence and preparedness to assist victims. Survey items were adapted from existing trafficking-awareness literature and prior educational research to ensure relevance to a university student population.

All substantive items were measured using a six-point Likert-type scale ranging from strongly disagree (1) to strongly agree (6). This response format was selected to reduce neutral responding and encourage directional responses. Given the exploratory nature of the study and the conceptual heterogeneity of the constructs, items were analyzed individually rather than combined into composite scales. No composite scales or subscale scores were constructed. Reverse-oriented items were retained as individual items, identified in the tables, and interpreted according to their wording. Lower agreement with reverse-oriented items generally reflected more victim-centered or less paternalistic responses.

Awareness was assessed using items measuring participants' understanding of what constitutes human trafficking and their exposure to trafficking-related information, including knowledge of coercion, fraud, and restraint, as well as awareness of anti-trafficking organizations and media exposure. Attitudes and misconceptions were assessed using items related to beliefs about victim agency, choice, and ability to leave trafficking situations. Empathy was measured using items assessing emotional and affective responses toward trafficked individuals, including feelings of empathy, anger, concern, and emotional engagement. Perceived preparedness was assessed through items measuring participants' confidence in their ability to assist trafficked individuals, engage in advocacy, raise public awareness, and address structural barriers.

Participants also reported whether they had previously completed coursework that included human trafficking content. This variable was coded dichotomously as yes/no and used to examine associations between prior trafficking-related education and awareness, attitudes, empathy, and perceived preparedness.

Data Analysis

All analyses were conducted using SPSS Statistics. Descriptive statistics were calculated for all survey items. Bivariate relationships between demographic variables and survey responses were examined using Pearson correlation coefficients. Several demographic variables were dichotomized to ensure adequate group sizes, including academic status (undergraduate vs. graduate) and academic discipline (nursing/physician assistant vs. all other majors). Statistical significance was evaluated at $p < .05$. Academic status reflected undergraduate versus graduate standing; prior coursework indicated whether participants had completed a course including human trafficking content; and academic discipline was coded as nursing/physician assistant versus all other majors.

Because the study analyzed items individually rather than as composite scales, the analysis was designed to identify item-level patterns in awareness, misconceptions, empathy, and perceived preparedness rather than to validate a new scale or produce subscale scores.

Results

Participants demonstrated high agreement with core elements of human trafficking (Table 2). Most respondents agreed that trafficking involves kidnapping (91.1%), fraud (74.9%), and coercion (81.1%) within the sex industry, and that individuals engaged in prostitution may be trafficked if prevented from leaving (75.5%). Across all definitional items, endorsement of misconceptions was low, and no significant relationships were found between demographic variables and these responses.

Table 2.

Agreement with Statements Defining Human Trafficking (N = 179)

Statement	Agree (%)	Disagree (%)
Kidnapping in sex industry	91.1	8.9
Fraud in sex industry	74.9	25.1
Coercion in sex industry	81.1	18.9
Restrained from leaving prostitution	75.5	24.5

Note. Responses were measured on a six-point Likert scale ranging from strongly disagree to strongly agree.

Participants largely rejected narratives suggesting that trafficked individuals can freely leave or choose their circumstances (Table 3). Several reverse-oriented items reflect commonly held misconceptions regarding victim agency and choice. Most respondents disagreed that trafficked persons can leave their circumstances (86.6%), and most disagreed that women (88.3%) or girls (89.4%) choose to be trafficked. By contrast, 82.1% agreed that trafficking is not a matter of personal choice. Responses regarding police involvement were more mixed; although most participants disagreed that victims could go to police but choose not to (77.7%), approximately one-fifth expressed some agreement (20.2%). The belief that some women choose to be trafficked was positively correlated with membership in the nursing/physician assistant major group ($r = .171$, $p < .05$).

Table 3.

Responses Regarding Trafficked Persons' Agency, Choice, and Ability to Leave

Statement	Agree (%)	Disagree (%)
A trafficked person has the ability to leave her circumstances (R)	11.2	86.6
Some women choose to be trafficked (R)	9.0	88.3
Some girls choose to be trafficked (R)	7.8	89.4
A trafficked person could go to the police, but she chooses not to (R)	20.2	77.7
It is not a person's choice to be trafficked	82.1	15.6

Note. (R) indicates a reverse-oriented item. Percentages may not total 100 due to missing responses and rounding. N = 179.

Self-reported awareness of sex trafficking was high (Table 4). Most participants reported exposure through news media (75.4%), reading about trafficking (78.2%), and public service announcements (73.8%). Over half reported awareness of organizations working against trafficking (62.6%). Although most participants disagreed that they were not informed about trafficking (67.0%), a notable minority expressed some level of agreement (29.2%). Feeling uninformed was negatively associated with being White ($r = -.150$, $p < .05$) and with having taken a course that included trafficking content ($r = -.183$, $p < .05$).

Table 4.

Awareness of Human Trafficking (N = 179)

Statement	Agree (%)	Disagree (%)
Aware of anti-trafficking organizations	62.6	34.7
Heard about trafficking in the news	75.4	21.8
Read about sex trafficking	78.2	18.4
Seen public service announcements	73.8	23.5
Not informed about trafficking (R)	29.2	67.0

Note. (R) indicates a reverse-oriented item. Percentages may not total 100 due to missing responses.

Participants expressed mixed views regarding intervention and decision-making on behalf of trafficked individuals (Tables 5 and 6). Just over half (54.7%) agreed that outsiders should intervene despite a victim's objections, while fewer endorsed outsiders making decisions about daily living (37.4%). Support for intervention was significantly associated with being male and being an undergraduate student. Additionally, agreement that outsiders should make decisions for trafficked individuals was positively associated with having a physician assistant or nursing major ($p < .05$).

Table 5 presents responses to items assessing attitudes toward intervention on behalf of trafficked individuals. In these items, "outsiders" refers to people outside the trafficking situation who may attempt to intervene or provide assistance, such as healthcare providers, service providers, advocates, law enforcement officers, or other community members. "Objections" refers to objections expressed by the trafficked person, not by the trafficker or another third party. These items were included to assess whether respondents endorsed survivor-centered assistance or more paternalistic forms of intervention in which others act or make decisions despite the trafficked person's stated preferences.

Table 5.

Attitudes Toward Intervention (N = 179)

Statement	Agree (%)	Disagree (%)
Intervene despite objections (R)	54.7	31.8
Outsiders make decisions for victims (R)	41.9	45.2
Outsiders control daily living decisions (R)	37.4	53.6

Note. (R) indicates a reverse-oriented item. For these items, agreement reflects greater endorsement of paternalistic or non-autonomous intervention attitudes. Lower agreement generally reflects a more victim-centered response. Percentages may not total 100 due to missing responses.

As shown in Table 6, greater agreement with paternalistic intervention items was associated with sex, class standing, and major, with male students and undergraduates showing stronger agreement on selected items.

Table 6.

Correlations Between Intervention Attitudes and Demographics

Variable	Sex	Class	Major
Intervene despite objections (R)	−.239*	−.220*	—
Make decisions for victims (R)	—	−.207*	.193*
Control daily living (R)	−.220*	−.164*	—

Note. (R) indicates a reverse-oriented item. Sex coded as female = 1, male = 0. Class coded graduate = 1, undergraduate = 0. Major coded nursing/physician assistant = 1, other = 0. $p < .05$.

Participants reported strong emotional engagement with the issue of trafficking (Table 7). Most respondents agreed that they felt anger (91.6%) and empathy (90.5%) toward trafficked persons, while most disagreed with statements indicating indifference, including that trafficking does not upset them (85.5%) or that they do not care about trafficking (88.3%). Approximately 62.6% reported becoming emotional when thinking about trafficking. No significant demographic correlations emerged for emotional response variables.

Table 7.

Emotional Responses to Sex Trafficking (N = 179)

Statement	Agree (%)	Disagree (%)
Angry about trafficking	91.6	3.9
Empathize with trafficked persons	90.5	5.6
Trafficking does not upset me (R)	10.0	85.5
Do not care about trafficking (R)	7.2	88.3
Become emotional thinking about trafficking	62.6	28.5

Note. (R) indicates a reverse-oriented item. Percentages may not total 100 due to missing responses.

Participants generally expressed confidence in their ability to contribute to anti-trafficking efforts (Tables 8 and 9). Nearly 60% believed they could make a difference for trafficked individuals, and over half felt capable of raising public awareness or addressing structural inequalities. Belief in one's ability to make a difference was positively associated with graduate student status, prior coursework on trafficking, and having a physician assistant or nursing major ($p < .05$). Similar associations emerged for political engagement and addressing structural barriers. No demographic variables were significantly associated with perceived ability to raise public awareness.

Table 8.

Perceived Ability to Assist Trafficked Individuals (N = 179)

Statement	Agree (%)	Disagree (%)
Can make a difference	59.2	35.8
Political involvement	59.8	37.4

Raise public awareness	74.3	22.9
Address structural inequalities	58.7	30.1

Note. Percentages may not total 100 due to missing responses.

Table 9.

Correlations Between Perceived Ability to Assist and Demographics

Variable	Class	Course	Major
Can make a difference	.186*	.246*	.217*
Political involvement	—	.177*	.172*
Raise public awareness	—	—	—
Address structural inequalities	—	—	.166*

Note. Class coded graduate = 1, undergraduate = 0. Course coded yes = 1, no = 0. Major coded nursing/physician assistant = 1, other = 0. $p < .05$.

Discussion

The purpose of this study was to examine human trafficking awareness, attitudes and misconceptions, empathy toward trafficked individuals, and perceived preparedness among university students, with particular attention to students preparing for patient-facing healthcare roles. Guided by prior research emphasizing healthcare providers as frontline identifiers, the findings underscore the importance of trafficking-related education while also revealing persistent misconceptions that may interfere with effective identification and intervention.

Overall, participants demonstrated strong baseline knowledge of human trafficking definitions and largely rejected the notion that trafficking is a matter of personal choice. Most recognized coercion, fraud, and restraint as defining elements of trafficking, indicating that overt misconceptions were uncommon. However, this conceptual understanding did not extend uniformly to beliefs about victim behavior, particularly regarding why victims may not disclose exploitation or seek assistance. Mixed responses to items concerning police involvement suggest lingering uncertainty about victim agency and help-seeking, a pattern consistent with prior research documenting gaps between definitional knowledge and applied understanding (Corthay, 2019; Lorvinsky et al., 2023).

Contrary to some prior studies, gender was not significantly associated with empathy toward trafficked individuals. This divergence should be interpreted cautiously, as it may reflect sample characteristics or institutional context rather than broader trends. From a communication perspective, empathy functions not merely as an affective disposition but as a clinical competency that shapes how providers interpret patient narratives and respond to vulnerability. The absence of strong gender effects highlights the importance of formal education and professional socialization in cultivating trauma-informed communication rather than assuming empathy develops naturally across demographic groups.

Participants demonstrated high levels of emotional engagement with the issue of trafficking, including strong expressions of empathy and anger toward exploitation. These findings suggest that students were emotionally responsive to the issue, although emotional concern alone does not ensure appropriate intervention. The observed association between nursing/PA majors and the belief that some women choose to be trafficked suggests that trafficking-related knowledge may remain uneven even among students preparing for patient-facing roles. These findings highlight the need for education addressing coercion, trauma bonding, and dependency, as students may simultaneously reject overt victim-blaming narratives while endorsing subtler misconceptions that undermine identification.

Prior trafficking-related coursework appears to play an important role in awareness and perceived preparedness. Students who reported prior coursework addressing human trafficking were less likely to feel uninformed and more likely to express confidence in their ability to assist victims. These findings suggest that structured education may strengthen students' readiness to recognize and respond to trafficking-related concerns, aligning with existing literature identifying education as a key factor in preparedness among healthcare students (Donahue et al., 2019; Raker, 2023; Olivieri et al., 2024). Importantly, informal exposure through media and public discourse may be insufficient to correct misconceptions, reinforcing the need for structured, curriculum-based instruction.

Participants expressed ambivalence regarding intervention, particularly when it involved acting against a victim's stated preferences or making decisions on their behalf. While many endorsed the need for intervention, fewer supported direct control over daily living decisions, suggesting uncertainty about ethical boundaries and professional authority. This ambivalence is concerning given the well-documented "dark figure" of trafficking victims who interact with healthcare systems without being identified (Bales et al., 2019). Nurses and PAs are often among the professionals best positioned to recognize trafficking indicators, making clarity around ethical, trauma-informed intervention essential.

Implications for Nursing and PA Education

These findings highlight the need to integrate standardized human trafficking education into nursing and PA curricula. Effective training should move beyond definitional awareness to address misconceptions about victim behavior, trauma bonding, and help-seeking, while also providing clear guidance on ethical intervention and referral pathways. Embedding this content early in clinical training may help ensure that students enter practice with both cognitive competence and communicative readiness.

From a communication standpoint, training should emphasize trauma-informed dialogue, recognition of nonverbal and contextual indicators, and strategies for creating psychologically safe environments for disclosure. Such competencies are central to frontline identification and cannot be assumed to develop without intentional instruction.

Limitations and Future Research

Several limitations should be noted. The use of a convenience sample from a single institution limits generalizability, and self-reported measures may be influenced by social desirability. Additionally, the cross-sectional design precludes conclusions about causality. Future research should examine how specific instructional strategies, such as simulation-based training or interdisciplinary coursework, influence clinical decision-making and long-term practice behaviors. Longitudinal studies assessing whether educational exposure translates into sustained identification and intervention would further strengthen the evidence base.

Conclusion

This study demonstrates that university students reported strong baseline awareness of human trafficking but also exhibited persistent uncertainty regarding victim behavior and intervention. Prior trafficking-related coursework appears to play an important role in awareness and perceived preparedness, while gaps in applied understanding highlight the need for more comprehensive, trauma-informed education. By integrating structured human trafficking education into healthcare-related curricula, institutions can strengthen frontline communication, reduce underidentification, and improve pathways to support for trafficked individuals.

References

Bales, K., Hesketh, O., & Silverman, B. (2015). Modern slavery in the UK: How many victims? *Significance, 12*(3), 16–21. https://doi.org/10.1111/j.1740-9713.2015.00824.x

Bales, K., Murphy, L. T., & Silverman, B. W. (2019). How many trafficked people are there in Greater New Orleans? Lessons in measurement. *Journal of Human Trafficking, 6*(4), 375–387. https://doi.org/10.1080/23322705.2019.1634936

Barrick, K., & Pfeffer, R. (2021). Advances in measurement: A scoping review of prior human trafficking prevalence studies and recommendations for future research. *Journal of Human Trafficking, 10*(1), 1–19. https://doi.org/10.1080/23322705.2021.1984721

Butler, R., Monsalve, M., Thomas, G. W., Herman, T., Segre, A. M., Polgreen, P. M., & Suneja, M. (2018). Estimating time physicians and other health care workers spend with patients in an intensive care unit using a sensor network. *The American Journal of Medicine*, *131*(8), 972.e9–972.e15. https://doi.org/10.1016/j.amjmed.2018.03.015

Chisolm-Straker, M., Richardson, L. D., & Cossio, T. (2012). Combating slavery in the 21st century: The role of emergency medicine. *Journal of Health Care for the Poor and Underserved, 23*(3), 980–987. https://doi.org/10.1353/hpu.2012.0091

Cole, E. (2017). Action against human trafficking and slavery. *Nursing Standard, 31*(49), 26–28. https://doi.org/10.7748/ns.31.49.26.s23

Corthay, E. (2019). The identification of victims of human trafficking: Definition, challenges and states' obligations. *Indonesian Journal of International Law, 16*(2). https://doi.org/10.17304/ijil.vol16.2.752

Donahue, S., Schwien, M., & LaVallee, D. (2019). Educating emergency department staff on the identification and treatment of human trafficking victims. *Journal of Emergency Nursing, 45*(1), 16–23. https://doi.org/10.1016/j.jen.2018.03.021

Hansen, M. A., & Johansson, I. (2025). Gender differences in attitudes and perceptions of human trafficking: Are they driven by knowledge gaps? *Journal of Human Trafficking*, 1–22. https://doi.org/10.1080/23322705.2025.2539000

Herrero-Villoria, C., Picornell-Lucas, A., & Patino-Alonso, C. (2022). Attitudes toward sex trafficking of university students through a multivariate statistical approach. *Journal of Evidence-Based Social Work*, *19*(6), 730–745. https://doi.org/10.1080/26408066.2022.2107413

Lorvinsky, J., Pringle, J., Filion, F., & Gagnon, A. J. (2023). Sex trafficking survivors' experiences with the healthcare system during exploitation: A qualitative study. *PLOS ONE*, 18(8), e0290067. https://doi.org/10.1371/journal.pone.0290067

Olivieri, S. S., Diamond-Caravella, M. L., & Etcher, L. N. (2024). Closing the educational gap on human trafficking: There’s no better time than now! *Journal of Emergency Nursing, 50*(4), 551–566. https://doi.org/10.1016/j.jen.2024.01.007

Peck, J. L., & Meadows-Oliver, M. (2019). Human trafficking of children: Nurse practitioner knowledge, beliefs, and experience supporting the development of a practice guideline: Part one. *Journal of Pediatric Health Care*, *33*(5), 603–611. https://doi.org/10.1016/j.pedhc.2019.05.006

Quincy, B., Falteisek, K., & Johnston, J. (2020). Recognizing and responding to sex-trafficked minors in the healthcare setting. *Journal of the American Academy of Physician Assistants, 33*(9), 43-47. https://doi.org/10.1097/01.JAA.0000694972.27395.74

Raker, K. A. (2020). Human trafficking education: A guide for nurse educators. *Journal of Professional Nursing, 36*(6), 692–697. https://doi.org/10.1016/j.profnurs.2020.09.015

Raker, K. A. (2023). An examination of nurse educators’ knowledge, attitudes, instructional beliefs, and instructional practices of human trafficking. *Journal of Professional Nursing, 47*, 35–45. https://doi.org/10.1016/j.profnurs.2023.04.002

Rapoza, S. (2021). *Nursing perceptions and attitudes toward sex trafficking: Does victim gender matter?* [Doctoral dissertation, University of Northern Colorado]. UNCOpen. https://digscholarship.unco.edu/dissertations/752

Seymour, R. G., & Silverman, B. W. (2023). How can we estimate modern slavery globally? *CHANCE, 36*(4), 22–29. https://doi.org/10.1080/09332480.2023.2290950

United Nations Office on Drugs and Crime. (2024). *Global report on trafficking in persons 2024*. https://www.unodc.org/documents/data-and-analysis/glotip/2024/GLOTIP2024_BOOK.pdf

Strengthening Democracy: Women's Participation in Local Governance Structures in India

Leena Thacker Kumar, Ph.D.
University of Houston-Downtown

Abstract
It is widely accepted that women's participation in governance is crucial for inclusive, equitable and sustainable development. India's 73rd constitutional amendment institutionalized women's participation by mandating a quota of at least a third of the seats in village-level governance institutions (called Panchayati Raj) being reserved for women. While recognizing the significance, and the positive outcomes of the quota system, this paper argues that mere numbers, in the form of quotas, while helpful, are not sufficient to address the myriad challenges women face while seeking equitable participation in policy making. Using the state of Jharkhand, India as its case study, the paper argues that networking and capacity building efforts by both the government and private sector are significant and necessary complements to truly enable the political, social, and economic development of women.

Key Words: Panchayat Raj Institutions, women's political participation, local governance, Jharkhand, capacity building

Introduction

The 73rd amendment of the Constitution of India, enacted in 1993, was a crucial step toward decentralization of power away from politicians and toward the people at the grassroots level. Among other provisions of this amendment, article 243D guaranteed that women hold a minimum of one-third (33%) of electoral seats in local governance structures in rural areas, known as Panchayati Raj Institutions (PRIs). Considering that women account for almost half (48.4%) of the Indian population, ensuring their adequate representation is fundamental to fostering gender equality and strengthening democratic institutions (World Bank, 2024). According to the Global Gender Gap Report (2023), India is ranked among the countries with the highest percentage of elected women in local governance with 44.4% thereby making it a global leader in women's representation at local level. As of 2025, there were over 1.4 million women in local decision-making roles in these local organizations (Akashvani News, 2025).

The Indian model of local governance has two distinctive features which allows for women in some states "to outnumber men in Panchayat bodies … (and) occupy leadership positions such as S*arpanch (*village head), *Pradhan (*chairperson) at district level" (Singh et al., 2025, p. 2930). First, as stated earlier, the Indian constitution mandates reservation of seats for women at the local levels and second, the constitution gives States the power and flexibility to exceed the one-third minimum requirement resulting in several states increasing the percentage to fifty and thereby further increasing the opportunity for "critical mass" presence of women in local governance.

This study considers the actual impact of these legislated quotas on the political, social, and economic development of women. Given that 63% India's population lives in rural areas (World Bank, 2024) and that about half of India's population are women, participation in local governance at the village level, i.e., within the Panchayat Raj Institutions (PRIs), is of crucial significance. The study specifically focuses on the state of Jharkhand, which, among other unique attributes, is among the few Indian states where a full 50% of PRI seats are reserved for women.

Literature Review

Grounded in research on "quota systems" and "critical mass theory," the study highlights the significance and achievements of quota policies in Jharkhand but also takes a clear-eyed view of the challenges women continue to face. Research shows that quotas and enhanced representation of women can augment "participation of women in decision-making bodies, thereby amplifying their voice" (Barodawala et al., 2025, p.3314). It is crucial to deepening democracy and advancing gender equality. Women's participation in governance has been shown to lead to effective policy outcomes, including lower corruption, reduced intersocial conflict, a more inclusive labor force, as well as economic growth leading to a more progressive society (Kumar & Ghosh, 2024; Burnwal, 2026). Additionally, exposure to women leaders has a trickle-down effect in that it reduces gender bias, raising aspirations among girls and women and increases family willingness to educate their daughters. (Barodawala et al., 2025; Singh et al., 2025) While there is a growing recognition of the crucial links between local governance, development, and gender equality, women face persistent challenges in participating in their local governments. It is thus important to identify not just the benefits but also the challenges faced by women in translating their presence into power and the crucial need for capacity building efforts to address these challenges.

In Jharkhand, despite the institution of the quota system, social customs, traditions, and norms that have traditionally subjugated women continue to persist, reducing the impact women can have in local governance, and ultimately slowing the social and economic development of these communities. As such, quota systems alone cannot, and will not, address the challenges women face, and must be complemented by capacity building efforts led by both government and private enterprise, to have the desired impact.

According to critical mass theory, the presence of a "considerable minority" in local governance structures is important to impact policy. The presence of more women, for instance, can affect deliberative processes, agenda setting, and political norms, resulting in substantive policy change that in turn leads to the further empowerment of women (Childs & Krook, 2008). Quotas are often a tangible way to rapidly increase women's ability to participate in policy making. Women's influence grows when their numbers reach a meaningful threshold, i.e., a critical mass. Around 30% is often treated as a practical benchmark to reach critical mass (Kumar & Ghosh, 2024).

Research on quotas of women in governance suggests "that the introduction of quota policies is followed by increased legislative attention to women's interests, issues and priorities" (Clayton, 2021, p.240). Quota systems are more likely to draw attention to issues related to women's rights, public health, poverty alleviation, access to water and education (Clayton, 2021; Barodawala et al., 2025). Besides having an impact on policy formulation, quotas tend to increase the percentage of minority women in politics as well as diversify their class and professional backgrounds (Barnes & Holman, 2020). Women's inclusion reflects a society that is moving towards greater acceptance of diversity and away from gender bias.

Gender quotas have proved to be effective at increasing women's political representation. Countries with such quotas have a higher representation of women in local government (by 7 percentage points) than countries without them (Berevoescu & Ballington, 2021). By mandating that women occupy at least 33% of seats in the PRIs, the 73rd Constitutional amendment of India established a gender quota that exceeds the 30% threshold commonly associated with the Critical Mass theory. India has adopted gender quotas to address longstanding gender disparities and to overcome the social and economic barriers linked with women's political exclusion. However, the potential positive consequences of quota systems are not guaranteed. (Clayton, 2021; Kumari & Guha Sinha, 2026). Mere presence in numbers does not automatically translate into positive policy outcomes for women.

The Context of Jharkhand

Jharkhand is a relevant case study, as it possesses some unique features that make it a powerful 'laboratory" for understanding the role and impact of women in PRIs on policy outcomes. First, Jharkhand is a relatively new state, formed in 2000. In 2011, the Jharkhand Panchayat Act mandated a 50% quota reservation for women in PRIs, significantly higher than the 30% threshold required in critical mass theory and the 33% requirement stipulated by the 73rd amendment. This put Jharkhand at the forefront of the drive to achieve gender-balanced local governance. It has a very large rural population (75%) and significantly lower literacy rates for women (56%) than for men (76%); both these conditions make it important, as well as difficult, to adopt policies that encourage improvements in the political, social, and economic statuses of women. Jharkhand, moreover, has one of the highest tribal (*Adivasi)* populations in India, over 26%. Tribal women face exaggerated challenges to development (Chiring & Komow, 2025). A recent study by Mandal et al. (2026) reveals that women empowerment among tribal women is lower than upper caste in Jharkhand.

However, while tribal women face higher scrutiny and obstacles, upper-caste women also face challenges. Upper-caste women encounter "socially normalized forms of proxying that draw less overt contestation yet remain deeply restrictive. These differences shape both the conditions of entry and the feasibility of moving from symbolic representation to recognized leadership" (Kumar & Bakshi, 2026. Pp. 13) The combination of all these factors makes Jharkhand a unique and interesting case study where women, tribal or non-tribal, have had a positive impact yet continue to face steep obstacles.

Positive Outcomes of Women's Participation in PRIs

The stipulation of a quota system in Jharkhand has had an overall positive impact on the sheer number of women participating in PRIs. Women's representation in PRIs has not just exceeded the 33% quota mandated by the 73rd. amendment, but for several years has also exceeded the 50% mandated by the State Act. According to the Ministry of Panchayat Raj, women hold around 51.57% to 56.35% of total seats in recent years (MoPR, 2026). This, in turn, has had a positive effect in encouraging more tribal women to step out of their homes and participate in the community's affairs, as well as supporting their own progress toward sustainable development goals. Across Jharkhand, female *Mukhiyas* (heads of Panchayat) have moved beyond symbolic representation to become policy actors who have impacted policies relating to water management, education, and social protection. Women's participation in PRIs is proving to be a very successful model of governance to address long-ignored issues such as sustainable employment, equal pay, women's property rights, irrigation, and access to water supply and overall infrastructure (Azim, 2013; Burnwal, 2026)

Women's participation has had the added benefit of creating a space for community women to share their problems with female leaders and with each other. This opportunity for women to interact in public creates ties outside the home, which generate "new political networks with shared interests and capacity for collective action" (Prillaman, 2023, p. 395). In some cases, they have been able to renegotiate the division of labor in the home. At the larger community level, women's opinions are receiving greater respect. They now have increased influence on decisions relating to, for instance, age of marriage for children and access to education for girls. Such success breeds increased interest among women to try their hand at politics and women are winning even unreserved panchayat seats in significant numbers (Lakshman, 2026). In fact, women who initially were reluctant to run for elections are now eager to remain in decision making roles (Jayal, 2006).

One striking example of the role of women in PRIs in Jharkhand is the case of Sita Sardar, a tribal *Mukhiya* from Ulidih village in Tatanagar. Sita overcame a series of barriers to become a leader with true political impact on her village. She formed the village Water, Sanitation, and Hygiene (WASH) committee. She led awareness campaigns, mobilized the community, and helped craft the village action plan to introduce systems for water quality testing. Today, every household in Ulidih receives water. "Jharkhand remains the only Indian state with a structured government-backed female WASH workforce. These women, like Sita, once humble water carriers, are now changemakers. The forest that once echoed with footsteps at dawn now resonates with leadership and hope" (Alang, 2025, *But That Has Changed Section).* This is a testament to the ripple effect of grassroots representation.

Villages led by women saw a noticeable rise in female engagement and a greater level of attentiveness to issues relating to women's policies (Accountability Initiative, 2010). Several studies (Beaman et al., 2010; Burnwal, 2026; Chattopadhyay & Duflo, 2004; Singh et al., 2025) show the benefits of women's participation in PRIs, including greater investments in services like water infrastructure, sanitation, health care, education, roads, and irrigation facilities.

Persistent Challenges to Women's Participation

Despite these achievements, women, particularly tribal women, face several challenges in the social, economic, and political arenas, which are often intertwined. Entrenched patriarchal norms inhibit women's participation. Many elected women continue to serve as a rubber stamp for their male family members, often referred to as a *sarpanch pati* (Burnwal, 2026; Singh et al., 2025). Family responsibilities fall almost entirely on women's shoulders, making it difficult to maintain the balance between household chores and PRI work. Women are expected to focus on every part of housework, from taking care of elders, to cooking, to raising and caring for children. Consequently, when they enter local governance, their household responsibilities often become prohibitive (Datta, 2023). Research finds that women often have no choice but to align themselves with the interests of the men in their households in fear of judgement from the outside and coercion and threats of domestic violence from the inside (Prillaman, 2023).

Lack of education or limited access to education for women from marginalized socio-economic background compared to their male counterparts further contributes to their lack of knowledge and power in running PRIs (Kumar & Ghosh, 2024; Singh et al., 2025). Lower rates of education reduce women's confidence in their work. In many cases, by the time they earn the skills to work on the council effectively, their term is over (Datta, 2023). Their inability to read training materials compromises their ability to understand their roles and responsibilities. They are also unable to read government documents or understand the budget, which makes them feel handicapped and limits their participation in discussions or potentially hinders their autonomy as they have to seek assistance from male members. As such, women are sometimes unaware of their political rights and responsibilities (Pandit, 2010). In addition, many women are not well equipped for their role because they have not interacted openly with men prior to their roles on PRIs (Palanithuria, 2002).

Compounding the lack or limited education women is the challenge of lower digital literacy rates among women which creates a major hurdle for performing their administrative functions efficiently. Women in rural India are deeply affected by the digital divide as the Ministry of Panchayati Raj transitions into digital governance platforms like eGramSwaraj. This digital divide impedes access to education, enhanced livelihood opportunities, and full participation in civic life (Mukherjee & Rakesh, 2025) which in turn compromises the economic independence of women (Dwibedi & Aptaprava, 2023).

Another major impediment for women is the Two-Child norm, which has been established as a criterion for participating in elections in several states. Candidates with more than two children are barred from running in PRI elections. Studies have shown that this policy disproportionately impacts women. Men may abandon their wives, hide the birth of their children, or give children up for adoption in order to run for PRI elections (Dwibedi & Aptaprava, 2023).

Surveys also reveal that women leaders receive less favorable assessments from their constituents than men, which indicates that women's work is not seen as valuable (Kumar & Ghosh, 2024). Evidence shows that a majority of women representatives do not run for reelection and, even when they do, often do not get reelected, and therefore cannot extend the administrative experience gained during one term to the next.

In addition to these social and economic challenges, women representatives face more direct political interference, including overbearing attitudes from Members of Parliament and members of legislative assemblies. Institutional practices within the PRIs sometimes act as an obstacle to women's ability to influence policy making. For example, relevant information essential for effective governance is purposefully blocked from women. Government information is sometimes only shared amongst fellow officials and routed through committees controlled by men resulting in exclusion of women from the decision-making loop (Singh et al., 2025). One panchayat Pradhan testified that "The local leadership of the party began to interfere with my functions and insisted that I should accommodate their interests, even if it came in violation of rules. They held local party meetings without my knowledge and shared the funds coming from different sources among themselves" (Datta, 1995, p. 836).

Most first-term women representatives, due to limited exposure to public life, find it difficult to deal with the higher forms of government bodies (Kumar & Ghosh, 2024). Female elected representatives also often encounter the negative attitudes of male officials, told repeatedly by these men feel that elected women have their positions only because of the quota system, that they lack their own ideas, and that they are easily duped by whatever advice existing officials feed them (Palanithurai, 2002). There is also the general perception, rooted in patriarchy, that women are simply not fit to be in positions of governance (Datta, 2023).

Despite these challenges, the combination of a quota system with a critical mass of women's representatives has, overall, led to a perceivable increase in the awareness, knowledge, and confidence levels of women, as decentralization and devolution have slowly taken root in the country. The fact that women have won 56% of the seats in 2022 in Jharkhand, more than the mandated reserved seats, reflects that voters have confidence in women leaders and women, who used to shy away from politics, have gained some confidence in running for elections from seats that are not reserved for women. "Once Dalit (*Adivasi)* women across Jharkhand gained legitimacy for public roles and political participation, they seldom retreated into the shadows" (Kumar & Bakshi, 2026, p.50). This burgeoning confidence and deepening of democratic processes is exemplified by the words of a women panchayat head (Mukhiya) of Garhwa village in Jharkhand where she states that, "Once we have stepped outside the home, we will never stop going out, not even until our last breath. It is a man's seat by default but if it becomes reserved for women then we will claim the seat. We will support women —what does matter if the seat is reserved or not? With women, we will remain active; we will work, no matter what. We are involved in other areas as well…And anyway, even if we accept that we lost the election, who will look after the people? It will still be us, won't it?" (Kumar & Bakshi, 2026, p.50)

Capacity Building as a Complement to Quotas: Implications for Policy and Practice

In analyzing the experiences of women in PRIs, it is clear that quota policies have had significant success. However, these policies alone cannot address the myriad challenges faced by women in PRIs. The quota policies need to be complemented with capacity building efforts, by both the government and private sector, targeted at many of the challenges faced by women. These capacity building efforts take many forms, including awareness and capability building in the form of sustained leadership training, legal awareness workshops, and literacy programs focused on digital and financial literacy. They also focus on strengthening self-help groups (SHGs), peer mentoring circles, and village-level women's collectives, all of which create alternative spaces where women develop confidence, discuss issues freely, and strategize solutions. Research shows a positive link between SHGs and economic empowerment of women (Burnwal, 2026).

The capacity building and training of PRIs has been one of the major goals of the Ministry of Panchayati Raj of the government of India. The Ministry provides technical, institutional, financial and programmatic support to strengthen PRIs. For example, the Jharkhand government has created the position of – the *Jal Sahiya* – to place tribal women at the forefront of water, sanitation, and hygiene efforts. These women (29,000 total) supported by the Drinking Water and Sanitation Department and UNICEF, operate across the state where “they are reshaping systems and social norms, and changing behaviors in their communities” (Alang, 2025, *But That Has Changed Section).* Involvement in these initiatives has had a ripple effect on women, who have sought positions within the PRI to continue this work.

Another example of a government led initiative is the Pradhan Mantri Gramin Digital Saksharta Abhiyan. This initiative is a digital literacy scheme by the Ministry of Electronics and Information Technology which focusses on teaching basic skills like operating computers and other digital devices, using email, browsing the internet and processing digital payments. As part of this initiative, the women from the Santhal tribe in Jharkhand utilize digital platforms to promote and sell their traditional arts globally (Mukherjee & Rakesh, 2025). This has in turn created a flywheel effect, promoting economic independence and higher levels of self-confidence, which should lead to further increased political participation.

Private organizations are also major contributors to capacity building. An example is the role played by the Corporate Social Responsibility (CSR) Group of the multi-billion-dollar Tata Steel Company. It actively empowers women, including panchayat leaders, through programs in digital literacy, self-help, and leadership, aiming to bridge gender gaps in governance. They also focus on equipping women to lead on community issues like health, sanitation, and economic development, often partnering with other global organizations to build capacity and create safer, more responsive villages. (United They Save, Tata Group 2021).

Tata Steel's CSR group partnered with Google to launch the internet "Saathi" ("friend") program, which provides smartphones and internet access to women in multiple states, including Jharkhand. The Saathis are rural women, originally with only basic literacy skills, who are trained to educate fellow women on how to use the internet. Each Saathi is tasked with training at least 600 women from other villages. The program has trained over 69,000 village women who, in turn, have taught around 25 million rural women. Internet Saathi has made government policies and news more accessible for women across the rural landscape. It has also created new opportunities for the Saathis themselves (Digital Goes Rustic, Tata Group 2020).

The CSR group has also worked with the *Sakhi Mandals* in Dumka district in Jharkhand to educate and create a platform for women to save money, access credit, and invest in income generating activities. This effort has not only improved the economic status of its members but also empowered them to take on leadership roles within their communities (Mukherjee & Rakesh, 2025).

Research shows that women who join self-help groups (SHGs) come to realize the importance of their own agency. They are more likely to participate in politics, including attending village assembly meetings and making demands of their leaders. SHGs increased women's autonomy. These SHGs, "bring together women with shared interests and foster political discussion and solidarity among women, in addition to developing civic skills through deliberation and information sharing" (Prillman, 2023, p. 392). Capacity building efforts, by both governments and private organizations, are invaluable in helping women overcome some of the challenges they face and serve as a necessary complement to the quota policies.

Conclusion

The 73rd amendment to the constitution of India aimed to increase women's participation in local PRIs by establishing quotas and thereby giving women the opportunity to have more agency and voice. Jharkhand, one of the newest Indian states, predominantly rural, with a large tribal population and lower literacy rates for women, went above the requirements of the Indian constitution by stipulating that 50% of the PRI seats be reserved for women. The quotas in

Jharkhand have undoubtedly had a positive impact, not only on the sheer number of women participating in PRIs, but also on the political influence and overall social and economic development of women. Yet, the positive impacts of the quota policy are not without social, political, and economic challenges, as many women continue to be impacted by traditional gender stereotypes and norms. To address these challenges, the quota policy alone is not sufficient. It must be complemented with capacity building efforts, directed at enabling women to form their own support groups, to become better able to advocate for themselves in larger numbers, and bring attention to issues that directly impact their wellbeing. These capacity building efforts, through government and private organizations, have produced many success stories already in Jharkhand.

Efforts to strengthen women's political leadership must go beyond legislated quota policies and address the broader social, economic, and institutional environments that shape and impact women's participation in practice. The strengthening of democracy not only depends on increasing women's representation in elected office but also in ensuring that capacity building efforts by both government and private organizations enable them to translate representation into lasting influence in practice. Great changes can come about to communities when women are given the space to be active participants in governing their communities and where institutional and social support are strong.

References

Accountability Initiative : Research and Initiative for Governance Accountability. (2010). *Decentralization: The Path to Inclusive Governance*.

Akashvani News (2025, March 4). *Over 1.4 million Women associated with Panchayat Raj Institutions: Union Minister Rajiv Ranjan Singh* [Press release].

Alang, A. (2025, June 17). *Lighting the way: How rural women are redefining leadership at the forefront of water, sanitation and hygiene in Jharkhand* [Press release]. UNICEF.

Azim S. (2013, September 19). TBI Women: After 30 years, Jharkhand Gets a Tase of Women Power. *The Better India.*

Barnes, T., & Holman, M. R. (2020). Gender Quotas, Women's Representation and Legislative Diversity. *Politics*, *82*(4), 1271–1286.

Barodawala, T., Joshi, P.D., & Singh V. (2025). A Catalyst for Change: Impact of Panchayat Raj on Women's Empowerment in India. *International Journal of Research and Innovation in Social Science,* IX (IV), 3310-3318.

Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2010). Political reservation and substantive representation: Evidence from Indian village councils. *India Policy Forum, 7,* 159-191.

Berevoescu, I., & Ballington, J. (2021). *Women's Participation in Local Government: A Global Analysis.* Governance and Participation Section, UN Working Paper.

Burnwal, R. K (2026). Role of Panchayati Raj Institutions in Women's Empowerment in Jharkhand: A Critical Analysis. *International Journal of Creative Research Thoughts,* 14 (6), 943-946.

Chattopadhyay, R., & Duflo, E. (2004). Women as policy makes: Evidence from randomized policy experiment in India. *Econometrica, 72(5),* 1409-1443.

Childs, S., & Krook, M.L. (2008). Critical Mass Theory and Women's Political Representation. *Political Studies, 56*, 725-736.

Chring. C.A., & Komow, B. (2025). Tribal Women in Local Governance: Exploring Socio-Economic and Cultural Dimensions of Participation in PRIs in Namsai District for Advancing Gender Equality. *Social Science Journal of Advanced Research, 5(5),* 1-8.

Clayton, A. (2021). How do Electoral Gender Quotas Affect Policy. *Annual Review of Political Science, 24,* 235-252.

Datta, P.K. (1995). Women and Panchayat Bodies in India. *The Indian Journal of Political Science, 56 (1/4),* 66-77.

Datta, P.K. (2023). Gender Quota and Women's Participation in Rural Local Bodies in India. *Indian Journal of Public Administration, 69(4),* 832-844.

Dwibedi, S.K., & Aptaprava, L. (2023). Representation of Tribal Women in Panchayat Raj Institutions. *Asian Journal of Management, 14(1),* 15-18.

Jayal, N,G. (2006). Engendering Local Democracy: The Impact of Quota for Women in India's Panchayats. *Democratization, 13 (1),* 15-35.

Kumar. R & Bakshi, E. (2026). *From margins to leadership: caste and gender dynamics in India's local governance.* Policy Development and Advisory Group, Align Report.

Kumar, S., & Ghost, A.K. (2024). *Elected Women Representatives in Local Rural Governments in India: Assessing the Impact and Challenges*, *No. 425*, Observer Research Foundation.

Kumari, T., & Guha Sinha, C. (2026). Democratic Decentralization and the Politics of Inclusion: Panchayat Empowerment in Jharkhand's Tribal Area. International Journal of Sociology and Humanities, *8(1),* 39-46.

Lakshman, A. (2026, April 26). Government Claims Women Winning Unreserved Panchayat Seats in 'significant numbers'. *The Hindu*. https://www.thehindu.com/news/national/govt-claims-women-winning-unreserved panchayat-seats-in-significant-numbers/article70893003.ece

Mandal, S., Banu, N., & Sarkar, R. (2026). *Empowerment Disparities between Tribal and Non-tribal Women in India: A Comparative Analysis.* Contemporary Voice of Dalit. 1-24.

Ministry of Panchayat Raj (2026). *Basic Statistics of Panchayat Raj Institutions.* Government of India.

Mukherjee, S., & Rakesh, K. M. (2025). Role of Digital Learning for Women in Tribal Societies of Jharkhand. *International Journal of Multidisciplinary Research*, *7*(2), 1–9.

Palanithuria, G. (2002). Impediments to Empowerment of Women: Experiences of elected women in Tamil Nadu. *The Indian Journal of Political Science*, *63*(1), 37–49.

Pandit, L. (2010). Political Leadership of Women: Constraints and Challenges. *The Indian Journal of Political Science*, *71*(4), 1130–1148. http://www.jstor.org/stable/42748942

Prillaman, S. A. (2023). Strength in numbers: How women's groups close India's Political Gender Gap. *American Journal of Political Science*, *67*(2), 390–410.

Singh, A, Maggo, N., Gupta, C.K., & Tandon, S. (2025). Gender and Grassroots Governance: Women's Political Participation in Local Self-Government in India in Comparison with Neighbouring Countries. *Journal of Local Self-Government, 23(S4),* 2925-2940.

Tata Group (2020) Digital Goes Rustic. https://www.tata.com/newsroom/community/tata-trusts internet-saathi-digital-goes-rustic

Tata Group. (2021). *United They Save.* https://www.tata.com/newsroom/community/tata-trusts-women-shgs

UN Women-Asia-Pacific. (2011). *To Become an Outstanding Panchayat Leader*. https://asiapacific.unwomen.otg/en/news-and-events/stories/2011/1/to-become-an-outstanding-panchayat-leader

World Bank Data Portal. (2024). *Indian Population section*. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=IN

World Economic Forum. (2023). *Global Gender Gap* Report. http://www.weforum.org/publications/global-gender-gap-report-2023/

Enteropathic Arthritis: A Systematic Literature Review of Prevalence and Treatment Options

Ava Smith

Dr. Praphul Joshi

Abstract

In this systematic literature review, there was a total of twenty resources used in order to understand enteropathic arthritis. These sources were picked from several different search terms including, "enteropathic arthritis + treatment", "enteropathic arthritis", and "seronegative spondyloarthropathy." All sources were found using PubMed, Google Scholar, Mayo Clinic, or Cleveland Clinic. In this systematic literature review, it is now clearer than ever that enteropathic arthritis is not studied enough. In many of these studies and journals, enteropathic arthritis was not being studied, but rather seronegative spondyloarthropathy as a whole, or ankylosing spondyloarthritis. However, in the research that was done, there were many findings when looking at enteropathic arthritis. This condition is an autoimmune condition that is considered a seronegative spondyloarthropathy. It is characterized with gastrointestinal symptoms as well as arthritic symptoms. This condition typically is associated with irritable bowel syndrome, which includes ulcerative colitis and Crohn's disease, it is also associated with other gastrointestinal conditions including celiac disease. There are two main types of enteropathic arthritis, axial and peripheral. Axial arthritis is characterized by pain in the spine, spondylitis, and sacroiliitis. Peripheral arthritis impacts the joints rather than the axial skeleton. When looking at the peripheral category, providers can further categorize the type of enteropathic arthritis by understanding how many joints are impacted. Providers must look at all symptoms and medical history in order to make a diagnosis and create a proper treatment plan. Treatment is different for every patient but there are options including medication, physical therapy, or acupuncture.

Key Words: Enteropathic arthritis, autoimmune, chronic

Introduction

Enteropathic arthritis has been studied extensively over the years, including studies on signs, symptoms, prevalence, and treatment. However, it still takes several years for a patient toreceive a diagnosis. There are several reasons why it takes so long to have a diagnosis. The main cause is that enteropathic arthritis has symptoms that are extremely similar to other diseases, including other autoimmune conditions. Due to the difficulty of diagnosis, this causes years where patients are suffering with symptoms that are not being treated or not being treated properly.

Methods

Using the search terms "enteropathic arthritis + treatment", "enteropathic arthritis", "axial spondyloarthritis" and "seronegative spondyloarthropathy", and filtering literature that was not before 2021 with the exception of one article from 2002, a total of 4,387 results were found.

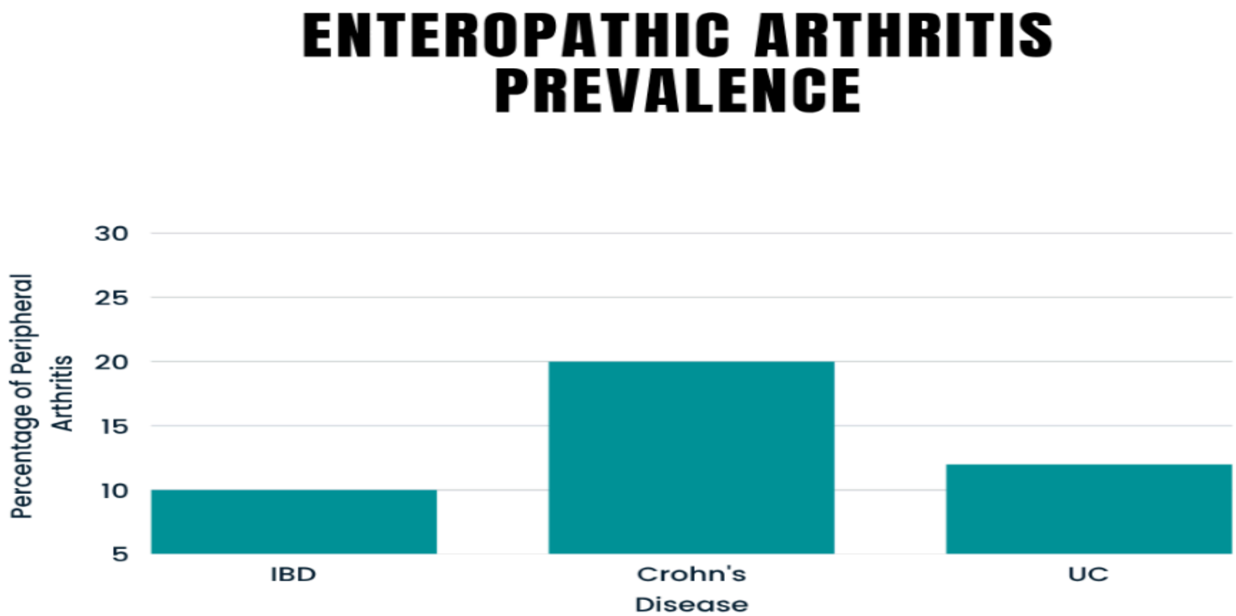
With these results, resources were chosen based on relevance of information for the systematic literature review and studies that were done in each journal. PubMed, Google Scholar, Mayo Clinic, and Cleveland Clinic were the sources used to find journals and reputable articles.

Results

Before understanding enteropathic arthritis, it is important to understand what type of condition it is. Enteropathic arthritis, or EnA, is a seronegative spondyloarthropathy (Sen & Hurley, 2023). Seronegative spondyloarthropathies are a group of autoimmune disorders that affect the joints. Subtypes include psoriatic arthritis and arthritis due to IBD (Sen & Hurley, 2023). These disorders can be hard to diagnose as they are extremely rare throughout the world, with only 0.5%-1.9% of the world's population diagnosed to have them (Sen & Hurley, 2023). However, the one symptom that allows providers to diagnose a seronegative spondyloarthropathy condition they can look for is enthesitis. Enthesitis is when the site where ligaments and tendons connect with the bone is inflamed (Sen & Hurley, 2023). This symptom is unique to these conditions (Sen & Hurley, 2023). When enthesitis is observed on imaging, providers can also look for a specific gene that is found in a large percentage of patients. This gene is the HLA-B27 gene (Sen et al., 2023). HLA-B27 is a type of human leukocyte antigen (McMichael & Bowness, 2002). A human leukocyte antigen helps the immune system identify cells as foreign or natural in the body (Nordquist & Jamil, 2023). One study demonstrated that 94% of patients suffering from ankylosing spondylitis, a type of seronegative spondyloarthropathy, have the HLA-B27 gene (McMichael & Bowness, 2002). How the gene is connected to these conditions is unclear. However, one theory states that HLA-B27 binds to a certain peptide which then the body's T-cells recognize this as foreign, causing the immune system to attack (McMichael & Bowness, 2002). In the case of spondyloarthropathies, it explains that these peptides that are bonded to HLA-B27 are typically found in the joints and other areas of symptoms of this type of arthritis (McMichael & Bowness, 2002). However, it is important to remember that although this theory has sense to it, it is only an example of some of the theories that have circulated on the importance of HLA-B27 and its significance to spondyloarthropathies.

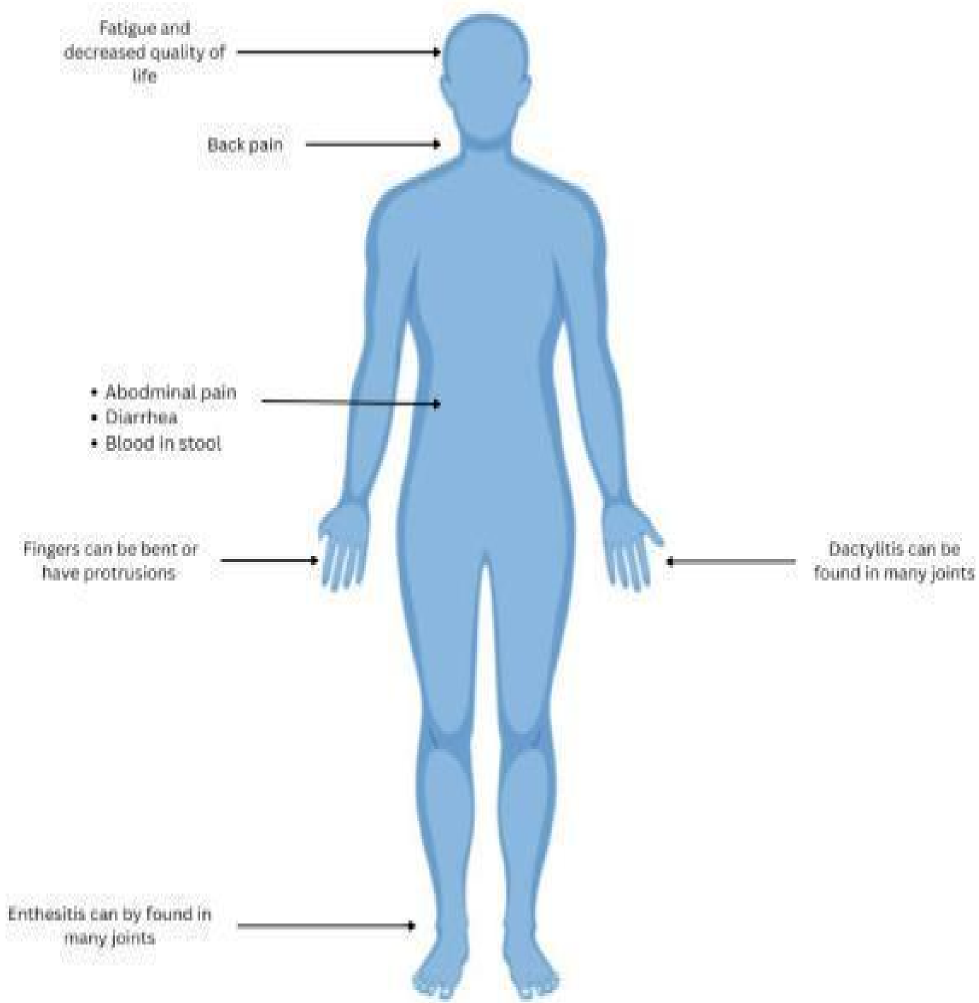
EnA is an inflammatory arthritis that is typically associated with gastrointestinal conditions. It can be caused by inflammatory bowel disease (IBD), Crohn's, or ulcerative colitis (UC) (Professional, 2025). EnA is chronic and cannot be cured, but there are ways to treat the symptoms which will be discussed later (Professional, 2025). This condition causes severe pain in the joints as well as damage to those joints (Professional, 2025). There are several different ways to categorize the types of enteropathic arthritis. One type is axial arthritis. This arthritis targets the spine as well as the patient's sacroiliac joints and causes stiffness typically in the morning. Axial is identified typically by sacroiliitis and spondylitis (Shahid & Brent, 2024). The other type of enteropathic arthritis is peripheral arthritis. This arthritis, unlike axial, targets the patient's joints in the appendicular skeleton. Therefore, it is classified by identifying enthesitis and dactylitis (Shahid & Brent, 2024). Peripheral arthritis can then be divided into two more categories; oligoarticular and polyarticular. Oligoarticular impacts less than five joints in the patient. The hips and knees will typically be affected by this type of arthritis. It is also more closely associated with a patient's gastrointestinal symptoms than polyarticular peripheral arthritis. Oligoarticular also leads to flares where symptoms get more severe. Polyarticular describes when five or more joints are impacted. Whereas oligoarticular targets larger joints, polyarticular targets smaller joints, including the hands. This type is not dependent on gastrointestinal symptoms and is usually long-lasting (Shahid & Brent, 2024). Enteropathic arthritis can be further classified into additional subtypes. These subtypes can be broken down further into three categories: Type 1, Type 2, and Type 3 (Shahid & Brent, 2024). Type 1 is oligoarticular where less than five joints are impacted, but it is associated with inflammation in the gastrointestinal system leading to flares. Type 2 is polyarticular where more than five joints are impacted and where both sides are affected (symmetric). This type is also chronic, meaning there is constant pain rather than flares that are seen in Type 1, and it does not depend on gastrointestinal symptoms. Type 3 is more of a combination between the first two and can indicate axial and peripheral symptoms (Shahid & Brent, 2024). These two types of classifications may have differences in names. However, the classification foundation is the same. There is axial and peripheral, then peripheral divides into oligoarticular and polyarticular.

One of the main reasons that EnA is extremely difficult to diagnose is due to the low prevalence of it. Researchers have to consider the prevalence of enteropathic arthritis side-by-side with the prevalence of IBD, Crohn's disease, and ulcerative colitis. For example, it is estimated that 6.7 million people have IBD in the world (Shahid & Brent, 2024). However, the prevalence of patients that have arthralgia with IBD, or joint stiffness, which is a common sign of enteropathic arthritis, is broadly estimated to be 6% to 46% (Shahid & Brent, 2024). The prevalence of enteropathic arthritis can also change based on the type of gastrointestinal disease that the patient is suffering from. Most researchers describe the prevalence of enteropathic arthritis based on the type of gastrointestinal condition. It is reported that 5% to 20% of people that have IBD can also suffer from peripheral arthritis, whereas 20% of patients with Crohn's and 12% of patients with ulcerative colitis suffer from peripheral arthritis (Shahid & Brent, 2024). It is difficult to find common prevalence among different articles. This can be due to many reasons, but the primary reason is that this condition is extremely difficult to diagnose, therefore making it difficult to quantify how many people are suffering from it.



Enteropathic arthritis can either present before or after the onset of gastrointestinal symptoms (Shahid & Brent, 2024). The broad range of symptomatology and symptom onset can also make it difficult to find the proper treating provider, especially in smaller communities.

Many patients can go for years without a diagnosis and without treatment. These years without treatment also cause damage to the body, making it more difficult for patients to recover after the start of treatment. The signs and symptoms of these diseases must be examined for full-body impacts, primarily gastrointestinal and arthritic (Shahid & Brent, 2024). However, as mentioned above, it can be more difficult to give the proper diagnosis because arthritis symptoms can appear either before or after gastrointestinal symptoms (Shahid & Brent, 2024). One diagnosis that is common for patients to receive is Rheumatoid Arthritis. This diagnosis is typically given to patients that have arthritis symptoms, but no gastrointestinal symptoms yet. However, there are a few distinctions between other inflammatory and autoimmune arthritis types and enteropathic arthritis. Enteropathic arthritis clearly includes gastrointestinal symptoms, but what are the differences between it and other arthritis types? Enteropathic arthritis is characterized by a certain amount of symptoms, these include chronic back pain in axial enteropathic arthritis, severe stiffness in the morning, and changes in weight (Shahid & Brent, 2024). When looking at the whole picture, there are certain symptoms that providers will look for. These include pain in limbs or spine, pain in the abdomen, stool containing blood, and diarrhea (Professional, 2025). It is important for providers to understand both the abdominal symptoms and joint symptoms to make a proper diagnosis. Not looking at all signs and symptoms as a whole can cause misdiagnosis and mistreatment. Signs and symptoms might also be different for different patients. Some might have severe damage to their knees or they may have severe damage to their hands. Signs and symptoms will also change person-to-person depending on whether the patient suffers from axial or peripheral. As mentioned above axial affects the back and the pelvis whereas the peripheral affects joints (Shahid & Brent, 2024). Depending on the type of enteropathic arthritis, providers can scan for enthesitis (inflammation where bone meets ligament), and dactylitis (swelling of the fingers), these symptoms are commonly seen in patients that suffer from enteropathic arthritis and patients who suffer from other spondyloarthropathies (Wang & Tsai, 2023). Providers can also look for weight loss or weight gain (Shahid & Brent, 2024). The differences between patients and their signs and symptoms demonstrates why it is extremely difficult to diagnose enteropathic arthritis.



The variability of these signs and symptoms can also affect the diagnosis based on when they occur. It is more difficult to diagnose enteropathic arthritis when a patient suffers from joint symptoms first. If a patient hasn't experienced gastrointestinal symptoms yet, providers are much more likely to make an incorrect diagnosis based only on the signs and symptoms related to the arthritis, leading to diagnoses such as rheumatoid arthritis (inflammatory arthritis) or osteoarthritis (noninflammatory arthritis). It is especially difficult for patients who are diagnosed with osteoarthritis because typically providers treat symptoms of osteoarthritis with medications that can worsen the gastrointestinal symptoms associated with enteropathic arthritis. This can also lead to a delay in proper testing and treatment. Patients who are diagnosed with rheumatoid arthritis may be at an additional disadvantage, as rheumatoid arthritis and enteropathic arthritis are extremely different pathologies and have vastly different treatment paths (Shahid & Brent, 2024).

As discussed above, diagnosis is extremely difficult for patients that suffer with enteropathic arthritis. However, there are ways that a provider can properly test for the condition, starting with a thorough patient history to understand the importance of all the signs and symptoms together and what picture they paint (Shahid & Brent, 2024). Providers should make note of all of the symptoms that were mentioned above as well as spondylitis (inflammation of the spine) and sacroiliitis (inflammation of the sacroiliac joints). One of the best tests to assess the inflammation is the Schober test, which is used to examine whether the patient has the proper range of motion in the lumbar spine. The provider can perform this test by measuring the patient's back while standing and bending over. If the patient does not have five centimeters or greater change while bending over, their lumbar spine does not have the range of motion it should (Shahid & Brent, 2024). The occiput-wall test examines the extent of the axial arthritis.

This test is performed by having the patient stand with their back entirely touching the wall. If the patient is unable to touch the back of their skull to the wall, it is considered atypical (Shahid & Brent, 2024).

Along with looking for specific signs and symptoms, providers must also run blood tests and imaging. A complete blood count (CBC), rheumatoid factor, erythrocyte sedimentation rate, anti-endomysial antibodies, anti-transglutaminase antibodies, and C-reactive protein are all blood tests that can aid in diagnosing enteropathic arthritis (Shahid & Brent, 2024). A CBC examines the main components of the blood. It counts erythrocytes, leukocytes, hematocrit, hemoglobin, and platelets (Staff, 2023). The erythrocyte sedimentation rate can be used to test for inflammation by measuring how long it takes for erythrocytes to reach the bottom of the test tube. If the erythrocyte sedimentation rate is high, this typically means there is systemic inflammation. Therefore, those with enteropathic arthritis typically have high sedimentation rates due to the inflammation in the joints (Professional, 2024a). The C-reactive protein blood test is also used to identify inflammation, which causes this protein to be released. As such, those with enteropathic arthritis will have high levels of the C-reactive protein (Professional, 2022). The tests discussed here are only a small representation of which laboratory tests providers can run when enteropathic arthritis is suspected. In addition to laboratory studies, providers can also perform colonoscopies and take biopsies of the intestines (Professional, 2025). Providers should also consider performing imaging of affected joints to assess the damage from the condition.

Imaging such as MRIs, ultrasounds, CTs, and radiographs can help assess joint damage and help providers choose proper treatment, thereby avoiding additional damage and helping to better manage the patient's pain (Shahid & Brent, 2024).

There are several options for treatment of enteropathic arthritis. Unfortunately, it can take trial and error to find the appropriate treatment for each patient. A patient can decide to use traditional treatment or alternative treatment. Traditional treatment includes medications, medical procedures, and physical therapy. When looking at medication, a patient can decide between synthetic disease-modifying antirheumatic drugs (Sdmards) and biologic disease-modifying antirheumatic drugs (Bdmards) (Nimmana & Patel, 2025). Sdmards are typically used as first-line treatment options when trying to reduce inflammation. They include drugs such as methotrexate and sulfasalazine. The drugs that are included in this category all have different mechanisms of action to reduce the inflammation. Methotrexate reduces inflammation by inhibiting immune activity and inhibits the production of cytokines. Sulfasalazine focuses on decreasing the release rather than the production of cytokines (Nimmana & Patel, 2025). Another type of medication that patients can try are biologic disease-modifying antirheumatic drugs (Bdmards). These drugs work by inhibiting cytokines and certain aspects of the immune system.

The most common of Bdmards are tumor necrosis factor (TNF) inhibitors. These drugs work by inhibiting the tumor necrosis factor alpha (TNF-A). TNF-A is responsible for inflammation.

Some examples of TNF inhibitors include adalimumab (Humira), infliximab, and golimumab (Nimmana & Patel, 2025).

Along with medications, patients can have different medical procedures. One of the main options for patients is glucocorticoid steroid injections into the joint (Professional, 2025). These injections decrease inflammation, which then reduces the patient's pain (Hopkins). A provider can also perform arthrocentesis on the patient. During this procedure, the provider can insert a needle into the joint and then remove the excess fluid. This can reduce pain and swelling of the joint and is a good option for pain relief. The provider can also test this fluid to aid in a diagnosis (Professional, 2024b). As mentioned above, treatment may look different depending on the patient, and sometimes the proper treatment is a mix between medication, imaging, and medical procedures. The main goal of all treatment options is to reduce inflammation and swelling. If the inflammation is decreased, this can allow the patient to have a great amount of pain relief and can also help with joint damage and quality of life.

With the several different types of medications that are available, providers must also look at the effectiveness of each drug. Sulfasalazine demonstrated a 59% success rate in a group of 600 patients that suffered from spondyloarthropathy. Sulfasalazine is considered one of the most effective drugs in maintaining enteropathic arthritis symptoms (AlAmeel et al., 2022).

Methotrexate is also considered to be highly effective. One study showed the improvement of IBD symptoms while taking methotrexate. This study by the North American Crohn's Study Group demonstrated the effectiveness of methotrexate with Crohn's disease (CD) which is a form of IBD. In the group of 141 patients that suffered from CD 39% of patients went into remission when before the study, their symptoms were dependent on steroids (AlAmeel et al., 2022). It is unclear whether there are studies on the effectiveness of methotrexate when treating enteropathic arthritis. However, there are studies on the impact of the medication on rheumatoid arthritis.

Sulfasalazine and methotrexate are both Sdmards. Patients can use Sdmards or Bdmards or both if needed. The most common examples of Bdmards are adalimumab (Humira), infliximab, and golimumab (Nimmana & Patel, 2025). All three of these examples are TNF inhibitors. Adalimumab showed significant effectiveness in a study of 42 patients suffering from Crohn’s disease. This study showed that adalimumab was also effective in controlling symptoms of Crohn’s that were outside of the gastrointestinal tract, meaning it could be extremely useful in controlling enteropathic arthritis symptoms (Barkhodari et al., 2022). Infliximab was also successful in treating symptoms in a study of 24 patients that suffered from spondyloarthritis and Crohn’s disease. Golimumab was also effective in a study where 9 out of 12 patients with spondyloarthritis and Crohn’s disease had an improvement in gastrointestinal symptoms and arthritis symptoms (Barkhodari et al., 2022). Another common medication issue related to enteropathic arthritis diagnosis is finding the right medication for the patient. Each patient is different and therefore symptoms are different. Patients have varying degrees of condition severity – one patient may achieve pain and symptom relief with adalimumab while another patient must use methotrexate to achieve the same relief. In addition, due to the low amount of research performed on enteropathic arthritis, providers frequently only have access to studies on how medications are effective on spondyloarthropathies but not on enteropathic arthritis itself, which can be quite frustrating. Therefore, patients may look for other forms of treatment that do not involve medication or medical procedures.

If a patient wants to avoid medication and traditional treatment there are options for alternative treatment including heat compresses, acupuncture, meditation, and massage (Professional, 2025). Although these treatments may not be for everybody, they are good options for those who would like to avoid medication, or for those who have tried medication, and it has not been effective. Patients can also be encouraged to take medication or have steroid injections, while also participating in alternative treatment. Patients are also encouraged to participate in physical therapy to improve range of motion and strength (Brigham). Physical therapy is suggested to patients not only because of how it can increase movement and functionality, but it can also improve quality of life and the mental state of patients suffering with enteropathic arthritis, or other forms of spondyloarthritis (van et al., 2023).

There are also other forms of alternative treatment or mind-body practices that have shown to be extremely useful in not only decreasing pain but also improving quality of life, slowing activity of the condition, and improving movement (Wang et al., 2024). For instance, acupuncture is considered helpful in maintaining symptoms, especially in the joints. In a study done on mice, acupuncture was performed on six mice who had rheumatoid arthritis. This study demonstrated that the acupuncture lowered the joint inflammation in the mice (N & Y, 2024).

This study is reassuring because it shows that acupuncture can decrease inflammation, therefore decreasing pain. Although this study was done on mice, it plays a pivotal role in finding more treatment options for those with enteropathic arthritis or spondyloarthropathies.

Another alternative treatment for these diagnoses is meditation, which can be helpful in relieving symptoms. Patients with this condition not only deal with chronic pain, flares, and gastrointestinal symptoms. They also constantly struggle with the mental aspect of having an autoimmune condition. Most patients can still walk, but they have pain while doing so. Patients cannot enjoy the activities of daily life; they cannot walk around a grocery store without pain. Although the severity of this condition can vary vastly from patient to patient, it is important to understand that this condition greatly impacts the quality of life. With meditation, patients can begin to heal mentally (Sibilia et al., 2024). With a decreased quality of life, a patient’s mental health typically is impacted, which can actually increase a patient’s pain. The stress of having this condition can actually make the pain seem worse, therefore impacting mental health more and creating a vicious cycle. With meditation, patients can improve both their mental health and their physical symptoms by decreasing their stress levels (Sibilia et al., 2024).

An additional alternative treatment for spondyloarthropathy is massage. Massage is known for helping with pain, specifically musculoskeletal pain (Sibilia et al., 2024). There are several studies indicating that massage significantly decreased pain (Kocyigit et al., 2022).

Massage has helped patients in different parts of the world including a study in Sweden. The data from this study demonstrates that massage helps pain significantly; the article even mentions that there was a reduction of the desire for pain medication which is extremely reassuring (Kocyigit et al., 2022). Massage is not only beneficial for pain but also helps with relaxation and stress relief. As mentioned above stress relief can also decrease pain and ease the severity of the condition (Sibilia et al., 2024). These alternative treatment options demonstrate that symptoms can be treated with medications and injections, however, it is also extremely important to manage the patient's mental health.

Conclusion

With the use of PubMed, Google Scholar, Mayo Clinic, and Cleveland Clinic, a total of twenty sources were used. The systematic literature review gave these findings. Enteropathic arthritis is a rare autoimmune condition that affects everyday life. It is a type of seronegative spondyloarthropathy. It is characterized by severe pain (both chronic and in flares), gastrointestinal symptoms, and inflammation. Enteropathic arthritis is a type of inflammatory arthritis that is associated with gastrointestinal conditions including irritable bowel syndrome or inflammatory bowel disease, Crohn's disease, and ulcerative colitis. It is a condition that is not talked about nearly enough which leads to misdiagnosis and mistreatment. Even in research today, it is difficult to find studies on enteropathic arthritis by itself. The studies and literature reviews discuss spondyloarthropathies as a whole or discuss only IBD. This makes it harder for providers and communities to be more informed, which makes the symptoms less recognizable and the condition hard to diagnose. If providers and patients are more informed about this condition, it can lead to more treatment options and a greater support system for patients that are suffering from enteropathic arthritis.

References

AlAmeel, T., Al Sulais, E., & Raine, T. (2022). Methotrexate in inflammatory bowel disease: A primer for gastroenterologists. Saudi Journal of Gastroenterology, 28(4), 250–260. https://doi.org/10.4103/sjg.sjg_496_21

Barkhodari, A., Lee, K. E., Shen, M., Shen, B., & Yao, Q. (2022). Inflammatory bowel disease: Focus on enteropathic arthritis and therapy. Rheumatology and Immunology Research, 3(2), 69–76. https://doi.org/10.2478/rir-2022-0012

Brigham and Women's Hospital. (n.d.). Spondylitis treatment options. Spondylitis Treatment Options. https://www.brighamandwomens.org/medicine/rheumatology-inflammation-immunity/resources/spondylitis-treatment#:~:text=Spondylitis%20is%20a%20type%20of,bone%20followed%20by%20fus ed%20vertebrae.

Hopkins, J. (n.d.). Injectable corticosteroids | Johns Hopkins Medicine. Injectable Corticosteroids. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/injectable-corticosteroids

Kocyigit, B. F., Sagtaganov, Z., Yessirkepov, M., & Akyol, A. (2022). Assessment of complementary and alternative medicine methods in the management of ankylosing spondylitis, rheumatoid arthritis, and fibromyalgia syndrome. Rheumatology International, 43(4), 617–625. https://doi.org/10.1007/s00296-022-05267-1

McMichael, A., & Bowness, P. (2002). HLA-B27: natural function and pathogenic role in spondyloarthritis. Arthritis research, 4 Suppl 3(Suppl 3), S153–S158. https://doi.org/10.1186/ar571

N, W., & Y, L. (2024, September 1). [Effects of acupuncture on intestinal microbiota in mice with rheumatoid arthritis]. National Center for Biotechnology Information. https://pubmed.ncbi.nlm.nih.gov/39401831/

Nimmana, B. K., & Patel, P. (2025, December 13). Disease-modifying antirheumatic drugs (DMARDS). StatPearls [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK507863/

Nordquist, H., & Jamil, R. T. (2023, April 24). Biochemistry, HLA Antigens. StatPearls [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK546662/

Professional, C. C. medical. (2022, May 17). C-Reactive Protein (CRP) test: What it is, Purpose & Results. Cleveland Clinic. https://my.clevelandclinic.org/health/diagnostics/23056-c-reactive-protein-crp-test

Professional, C. C. medical. (2024a, February 20). What is an ESR test?. Cleveland Clinic. https://my.clevelandclinic.org/health/diagnostics/17747-sed-rate-erythrocyte-sedimentation-rate-or-esr-test

Professional, C. C. medical. (2024b, December 22). Should you have arthrocentesis?. Cleveland Clinic. https://my.clevelandclinic.org/health/treatments/14512-arthrocentesis-joint-aspiration

Professional, C. C. medical. (2025, March 21). What is enteropathic arthritis?. Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/23166-enteropathic-arthritis

Sen, R., Goyal, A., & Hurley, J. A. (2023, July 17). Seronegative spondyloarthropathy.

StatPearls [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK459356/

Shahid, Z., & Brent, L. H. (2024, April 19). Enteropathic arthritis. StatPearls [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK594239/

Sibilia, J., Berna, F., Bloch, J.-G., & Scherlinger, M. (2024). Mind-body practices in chronic inflammatory arthritis. Joint Bone Spine, 91(3), 105645. https://doi.org/10.1016/j.jbspin.2023.105645

Staff, M. C. (2023, January 14). Complete blood count (CBC). https://www.mayoclinic.org/tests-procedures/complete-blood-count/about/pac-20384919

van Wissen, M., Gademan, M., Vliet Vlieland, T., Straathof, B., Teuwen, M., Peter, W., van den Ende, C., & van Weely, S. (2023). Physical therapy in patients with rheumatoid arthritis and axial spondyloarthritis: The patients' perspective. Scandinavian Journal of Rheumatology, 52(6), 609–618. https://doi.org/10.1080/03009742.2023.2195726

Wang, C.-R., & Tsai, H.-W. (2023). Seronegative spondyloarthropathy-associated inflammatory bowel disease. *World Journal of Gastroenterology*, *29*(3), 450–468. https://doi.org/10.3748/wjg.v29.i3.450

Wang, J., Li, X., Yang, F., Guo, P., Ren, C., Duan, Z., & Zhang, Y. (2024). Efficacy and safety of mind-body exercise for patients with axial spondyloarthritis: A systematic review and meta-analysis. Journal of Orthopaedic Surgery and Research, 19(1). https://doi.org/10.1186/s13018-024-05072-5

Effects of an Educational Intervention on Energy Drink Use and Cardiovascular Awareness in College Students

Rachel Barker
Department of Public Health
Sam Houston State University

Praphul Joshi
Department of Public Health
Sam Houston State University

Abstract

Energy drink consumption is common among college students and may contribute to acute cardiovascular strain due to high caffeine content and stimulant combinations. This study evaluated the effects of a brief classroom educational intervention on energy drink knowledge, cardiovascular risk awareness, and self-reported behaviors among undergraduate students. Using a quasi-experimental pre-post design, students from 18 in-person classes completed anonymous paper surveys at baseline, immediately after a 15-minute educational session, and one month later, representing approximately 202 participants. Class-level analyses showed that mean knowledge scores increased from 8.06 to 9.44 out of 10 immediately after the session, representing a statistically significant improvement with a very large effect, $t(17) = 17.30$, $p < .001$, *Hedges' g = 3.89. At one month, knowledge declined to 8.26 and was not significantly different from baseline,* $t(17) = 0.55$, $p = .589$, *while post-intervention to follow-up scores decreased significantly,* $t(17) = 2.69$, $p = .016$. *Despite this decline, energy drink consumption frequency decreased significantly from baseline to follow-up, with a mean difference of −0.96,* $p = .045$, *while the reduction in amount consumed per occasion was not statistically significant. These findings demonstrate that a brief, low-resource classroom intervention can support meaningful reductions in energy drink use and can be incorporated into campus wellness programs, health courses, and student education initiatives, although reinforcement may be needed to sustain knowledge gains over time.*

Key Words: Energy drinks, college students, cardiovascular health, health education intervention, behavior change

Introduction

1.1 Background on Energy Drinks

Energy drinks are beverages marketed to improve alertness, concentration, and physical performance through caffeine and other stimulant ingredients. Although originally associated with athletic performance and endurance, they are now widely consumed by adolescents and young adults, especially college students (National Center for Complementary and Integrative Health [NCCIH], 2023). Their availability in convenience stores, vending machines, and campus retail locations has made energy drink use a common part of student life (Jefferson Health, 2023). College students often report using these products to manage fatigue, improve focus, and keep up with academic demands, but frequent reliance on stimulants may reinforce unhealthy patterns when used in place of sleep, hydration, and balanced nutrition (Jefferson Health, 2023; Mayo Clinic Staff, 2024).

The rapid rise in energy drink consumption is concerning because of its connection to cardiovascular health. Excessive caffeine intake and stimulant combinations found in energy drinks can increase heart rate and blood pressure, which may contribute to palpitations, hypertension, and arrhythmias, even among otherwise healthy individuals (Higgins et al., 2018; UC Davis Health, 2024). These effects may be especially concerning for students who consume energy drinks frequently, use them with other caffeinated products, or have underlying health conditions (Higgins et al., 2018). While up to 400 mg of caffeine per day is considered safe for most healthy adults, individual energy drinks often contain large amounts of caffeine, and some containers include multiple servings per can (Center for Science in the Public Interest [CSPI], 2024; Mayo Clinic Staff, 2024). As a result, students may unintentionally exceed recommended limits, especially when energy drinks are consumed alongside coffee, soda, or other caffeinated beverages (Mayo Clinic Staff, 2024).

In addition to caffeine, energy drinks often contain ingredients such as taurine, guarana, L carnitine, glucuronolactone, ginseng, B vitamins, sugar, and artificial sweeteners (Higgins et al., 2018). Guarana may contribute additional caffeine that is not always clearly recognized by consumers, while high sugar content can contribute to obesity and other cardiovascular risk factors (Ali et al., 2022; Higgins et al., 2018; UC Davis Health, 2024). Because many students may not fully understand these ingredients or their potential health effects, targeted education is needed to improve awareness and support healthier choices. This makes college campuses an important setting for brief educational interventions aimed at reducing risky stimulant use and promoting healthier energy management strategies.

1.2 Public Health Relevance

Energy drink consumption is a public health concern because it is common among adolescents and young adults and may contribute to preventable cardiovascular risk. Young adulthood is an important period for establishing long-term health behaviors, making college students a key population for early prevention efforts (CDC, 2024). Although energy drinks are often perceived as routine or harmless, excessive stimulant intake may increase cardiovascular strain, especially when combined with stress, sleep deprivation, or underlying health conditions (Higgins et al., 2018; UC Davis Health, 2024). This study is grounded in the Health Belief Model, which suggests that behavior change is influenced by perceived susceptibility, perceived severity, perceived benefits, perceived barriers, and cues to action (Alyafei & Easton Carr, 2024). Many college students may underestimate their personal risk from frequent energy drink consumption, which can reduce motivation to change behavior. By communicating cardiovascular risks, safe caffeine limits, and realistic alternatives, the educational intervention was designed to increase awareness and encourage healthier decision-making.

1.3 Purpose of the Study

The purpose of this study was to evaluate the short-term effects of a brief classroom-based educational intervention on college students' knowledge, awareness, and behaviors related to energy drink consumption and cardiovascular risk. Guided by the Health Belief Model, the intervention was designed to increase perceived susceptibility and severity, emphasize the benefits of behavior change, and provide practical strategies to reduce barriers to healthier choices. Although prior research has documented the cardiovascular effects of energy drinks, including increased heart rate and blood pressure, fewer studies have examined whether brief educational interventions can improve students' knowledge and influence consumption behaviors (Higgins et al., 2018; Martínez-Sánchez et al., 2023; UC Davis Health, 2024). This study addressed that gap by assessing baseline knowledge and behaviors, immediate learning and behavior-change intentions, knowledge retention, and reported behavior change one month after the intervention. Findings can inform scalable campus health education strategies designed to reduce risky stimulant use and promote cardiovascular health among college students.

1.4 Research Questions and Hypotheses

This study examined the effects of a brief classroom-based educational intervention on college students' knowledge and behaviors related to energy drink consumption. The following research questions and hypotheses guided the study:

RQ1: What were students' baseline levels of knowledge regarding energy drink ingredients, recommended caffeine limits, and cardiovascular risks, and what were their baseline energy drink consumption behaviors?

RQ2: Did mean class-level knowledge scores increase immediately following the educational intervention?

H1: Mean class-level knowledge scores would be significantly higher immediately after the intervention than at baseline.

RQ3: Were improvements in knowledge retained one month after the educational intervention?

H2a: Mean class-level knowledge scores at the one-month follow-up would be significantly higher than baseline scores.

H2b: Mean class-level knowledge scores at the one-month follow-up would be lower than immediate post-intervention scores.

RQ4: What knowledge gains and behavior-change intentions did students describe immediately following the educational intervention?

RQ5: Did class-level energy drink consumption frequency and amount consumed per occasion change from baseline to the one-month follow-up?

H3a: Mean class-level energy drink consumption frequency would be significantly lower at the one-month follow-up than at baseline.

H3b: Mean class-level energy drink consumption amount per occasion would be significantly lower at the one-month follow-up than at baseline.

RQ6: What health-related behavior changes and harm-reduction strategies did students report at the one-month follow-up?

1.5 Significance of the Study

Findings from this study can inform campus wellness initiatives by offering a scalable and low-resource intervention model that can be integrated into existing student health programming. Results can inform the development of workshops, classroom-based education, and peer-led initiatives aimed at reducing high-risk caffeine consumption. Incorporating evidence from this research into campus wellness strategies may strengthen prevention efforts, promote healthier beverage choices, and support broader cardiovascular health promotion goals within the university setting.

Methods

2.1 Study Design

This study utilized a quasi-experimental, repeated-measures pre-post design to evaluate the impact of a brief educational intervention on college students' knowledge, intentions, and behaviors related to energy drink consumption. Participants completed surveys at three time points: prior to the educational session (pre-intervention), immediately following the session (post-intervention), and approximately one month later (follow-up). This design allowed for the assessment of baseline knowledge and behaviors, immediate learning outcomes, and short-term retention and behavior change over time. Participation was voluntary, and students were allowed to skip any survey questions they did not wish to answer.

2.2 Participants

Participants were undergraduate students enrolled at Sam Houston State University (SHSU). Inclusion criteria required students to: (1) be currently enrolled as an undergraduate at SHSU, verified by enrollment in the course where the study was conducted; (2) attend a regularly scheduled in-person class in which the educational intervention was delivered; and (3) be able to provide informed consent, meaning they could read and understand English and voluntarily agree to participate. Exclusion criteria included: (1) individuals under 18 years old, as the study focused on adult undergraduate students; (2) non-English-speaking individuals, since all study materials, surveys, and presentations were conducted in English; (3) students not present during the in-class intervention and survey administration; and (4) individuals who did not provide informed consent, as participation was voluntary. A total of 18 undergraduate classes participated, representing approximately 202 students. The average number of participants per class was 11.19. Because the surveys were anonymous and individual responses could not be linked across the three data-collection periods, the class served as the unit of analysis.

2.3 Recruitment

Participants were recruited through course-based outreach at Sam Houston State University. The principal investigator (PI) contacted faculty across multiple academic departments to request permission to conduct the study during scheduled in-person class sessions. Courses that agreed to participate were scheduled for the educational intervention and survey administration. Students enrolled in these courses were invited to participate at the time of the class session. Participation was voluntary, and students were informed that they could decline participation or skip any survey questions without penalty. Demographic information, including age, sex, year in school, and major, was collected during the pre-survey. All responses were anonymous to ensure confidentiality.

2.4 Procedures

2.4.1 Pre-Survey

The pre-survey was administered at the beginning of the class session, prior to the educational presentation. It assessed participants' baseline knowledge of energy drink ingredients and cardiovascular risks, as well as baseline health behaviors, including energy drink consumption, physical activity, diet, and sleep. Demographic information was also collected. Surveys were paper-based, anonymous, and required approximately 5–10 minutes to complete.

2.4.2 Post-Survey

Immediately following the 15-minute educational session, participants completed a post-survey designed to measure changes in knowledge and intentions related to energy drink consumption. The knowledge questions were identical to those included in the pre-survey, allowing for direct comparison of baseline knowledge and immediate learning. Additional items assessed participants' intentions to modify behaviors, such as reducing energy drink consumption, selecting healthier alternatives, and improving sleep or dietary habits. The post-survey was paper-based, anonymous, and required approximately 5–10 minutes to complete.

2.4.3 One-Month Follow-Up Survey

Approximately one month after the initial session, participants completed a follow-up survey during a subsequent class session. This survey assessed retention of knowledge using the same knowledge questions as the pre- and post-surveys, as well as changes in health behaviors, including energy drink consumption, physical activity, diet, and sleep. Additional items evaluated self-reported behavior change related to the intervention and the use of harm-reduction strategies, such as reducing energy drink intake or choosing alternative beverages. The follow-up survey also included open-ended questions to capture qualitative information on lifestyle changes. The survey was paper-based, anonymous, and required approximately 10–15 minutes to complete.

2.5 Intervention/Education Program

The educational component consisted of a 15-minute classroom-based session focused on energy drinks and their potential effects on the cardiovascular system. The session covered recommended daily caffeine limits, common energy drink ingredients, and both short-term and long-term cardiovascular health risks associated with excessive consumption. In addition to risk awareness, the session emphasized practical strategies for reducing energy drink intake, including identifying hidden sources of caffeine, reading nutrition labels, and selecting healthier alternatives such as water, electrolyte drinks, or moderate caffeine options. Students were also encouraged to adopt sustainable energy management strategies, including adequate sleep, hydration, and balanced nutrition. The session was designed to be concise, interactive, and appropriate for a college student audience, with the goal of increasing awareness while promoting realistic and actionable behavior change.

2.6 Data Collection

Data were collected using anonymous paper surveys administered at three time points: before the intervention, immediately after the intervention, and approximately one month later. The same knowledge questions were included at each assessment to allow comparison of knowledge across time. Surveys also included demographic, behavioral, Likert-style, and open-ended items assessing energy drink use, health behaviors, learning, intentions, and reported lifestyle changes. All surveys were completed during scheduled class sessions. The pre- and post-intervention surveys required approximately 5–10 minutes to complete, while the follow-up survey required approximately 10–15 minutes. Completed surveys were collected immediately after administration and stored securely.

2.7 Data Management

All surveys were collected immediately after administration and stored securely to maintain participant confidentiality. Survey responses were entered into a password-protected dataset and aggregated at the class level for analysis using IBM SPSS Statistics. Data were checked for accuracy and completeness prior to analysis. All analyses were conducted using de-identified data, and results were reported in aggregate form.

2.8 Quantitative Data Analysis

Quantitative data were analyzed using IBM SPSS Statistics to examine changes in knowledge and health behaviors related to energy drink consumption at the class level. Individual survey responses were aggregated within each class to generate class-level means for all variables, and analyses were conducted using these aggregated data (N = 18 classes). Because surveys were anonymous and responses could not be linked for the same student across the pre-intervention, post-intervention, and follow-up assessments, individual-level repeated-measures analyses were not possible. Responses were therefore aggregated within each class, allowing class-level means and proportions to be compared consistently across the three data-collection periods. This approach also reflected the classroom-based delivery of the intervention, in which each participating class received the educational session as a group. Descriptive statistics, including means and standard deviations, were calculated to summarize class-level characteristics, baseline behaviors, and knowledge scores at pre-intervention, post-intervention, and one-month follow-up. Paired-samples t-tests were conducted to evaluate changes in knowledge scores across time points, including pre-intervention to post-intervention, pre-intervention to follow-up, and post-intervention to follow-up. Effect sizes were calculated using Hedges' g to assess the magnitude of observed changes. Additional paired-samples t-tests were conducted to examine changes in class-level health behaviors from pre-intervention to follow-up, including energy drink consumption frequency and amount, as well as sleep-related variables. These analyses were considered exploratory due to the relatively small number of classes. For knowledge items coded as binary variables, proportions of correct responses were calculated at the class level. Where appropriate, changes in these proportions across time points were examined descriptively due to limited variability in some items. Given the relatively small number of classes, results were interpreted with attention to both statistical significance and effect size to assess practical significance.

2.9 Qualitative Data Analysis

Qualitative responses from all open-ended survey questions were analyzed using an inductive thematic content analysis approach. Responses were reviewed systematically, and initial codes were developed directly from recurring ideas in the data rather than from predetermined categories. Similar codes were grouped into broader thematic categories representing knowledge gains, intended behavior changes, and reported lifestyle changes. Because participants sometimes expressed multiple ideas within a single response, responses could be assigned to more than one category. Categories were then reviewed for consistency, overlap, and clarity, and similar categories were combined when appropriate. Theme frequencies represent the number of times each theme was identified rather than the number of unique participants. Qualitative findings were summarized descriptively and used to complement and contextualize the quantitative results.

2.9.1 Immediate Learning Question

The immediate post-intervention survey included an open-ended question asking students whether they learned something new from the educational session and, if so, to describe what they learned. Identified concepts were grouped into thematic categories representing key learning points, including safe caffeine limits, cardiovascular and other negative health effects, ingredient awareness, safer alternatives to energy drinks, duration of stimulant effects, and energy drink intoxication.

2.9.2 Immediate Behavior Change Intentions

The immediate post-intervention survey included an open-ended question asking students whether they planned to change their behavior and to briefly describe any intended changes. Thematic categories captured intended actions such as reduction in energy drink intake, reduction in overall caffeine intake, avoidance of caffeine before exercise, substitution of energy drinks with alternative beverages, dietary improvements, increased attention to nutrition labels, and seeking medical evaluation for low energy. Additional categories reflected students reporting no planned change due to already low or controlled energy drink use, no energy drink consumption, or behavior changes that had occurred prior to the session.

2.9.3 One-Month Lifestyle Change

The one-month follow-up survey included an open-ended question asking students to describe lifestyle changes made over the past month to improve health and fitness. Identified themes included increased physical activity, reduced energy drink consumption, reduced overall caffeine intake, improved sleep habits, adoption of a more balanced diet, and increased attention to nutrition labels and ingredient content. Additional categories captured responses indicating no change, no prior energy drink use, or general positive change without specific detail. Less frequent but conceptually important themes such as reduced stress or anxiety, decreased headaches, and increased attention to existing health conditions were also recorded.

2.10 Ethical Considerations

Participation in the study was voluntary, and informed consent was obtained before data collection. Participants were informed that they could skip any survey questions or withdraw at any time without penalty. Surveys were anonymous, responses were stored securely, and all data were analyzed in deidentified aggregate form. The study protocol was reviewed and approved by the university's Institutional Review Board, and the research team completed required human subjects research training.

Results

3.1 Participant and Class Characteristics

At the class level, the mean student age was 21.39 years (*SD* = 0.66), and the mean proportion of male students was 0.14 (*SD* = 0.11), indicating that the sample was predominantly female. Students represented multiple undergraduate classifications and academic levels. Baseline descriptive characteristics, including knowledge scores, energy drink consumption, and sleep behaviors, are presented in Table 1.

3.2 RQ1: Baseline Knowledge and Energy Drink Consumption Behaviors

RQ1 examined students' baseline knowledge of energy drink ingredients, recommended caffeine limits, and cardiovascular risks, as well as baseline energy drink consumption behaviors. At baseline, the mean class-level knowledge score was 8.06 out of 10 (*SD* = 0.41), indicating that students answered approximately 81% of the knowledge questions correctly before the intervention. These findings suggest that baseline knowledge was generally high, although variation remained across classes in students' understanding of energy drink ingredients, caffeine limits, and cardiovascular risks. The mean class-level energy drink consumption frequency score was 2.47 (*SD* = 0.58), while the mean amount consumed per occasion was 1.62 (*SD* = 0.44). These results indicate that energy drink use was present across the participating classes, although consumption levels varied. Baseline descriptive findings are presented in Table 1.

Table 1
Class-Level Descriptive Characteristics at Baseline (N = 18 classes)

Variable	**Mean**	***SD***
Class size (students)	11.19	6.20
Mean age (years)	21.39	0.66
Proportion male	0.14	0.11
Baseline knowledge score (0–10)	8.06	0.41
Energy drink frequency	2.47	0.58
Energy drink amount per occasion	1.62	0.44
Sleep duration	2.14	0.49
Sleep quality	2.43	0.46

3.3 RQ2 and H1: Immediate Change in Knowledge

RQ2 examined whether mean class-level knowledge scores increased immediately following the educational intervention. H1 predicted that knowledge scores would be significantly higher immediately after the intervention than at baseline. Mean knowledge scores increased from 8.06 (SD = 0.41) at baseline to 9.44 immediately after the intervention. This represented a mean increase of 1.40 correct answers. A paired-samples t test indicated that the increase was statistically significant, $t(17) = 17.30$, $p < .001$. The effect was very large, Hedges' $g = 3.89$. Therefore, H1 was supported. The educational intervention produced a statistically significant and substantial immediate improvement in class-level knowledge. The results are presented in Table 2.

3.4 RQ3, H2a, and H2b: Knowledge Retention at One-Month Follow-Up

RQ3 examined whether improvements in knowledge were retained one month after the intervention. H2a predicted that follow-up knowledge scores would remain significantly higher than baseline scores. H2b predicted that follow-up knowledge scores would be lower than immediate post-intervention scores. At the one-month follow-up, the mean class-level knowledge score was 8.26 (SD = 1.75). Follow-up scores were not significantly different from baseline scores, $t(17) = 0.55$, $p = .589$, Hedges' $g = 0.12$. Therefore, H2a was not supported. Knowledge scores decreased significantly from the immediate post-intervention assessment to the one-month follow-up, $t(17) = 2.69$, $p = .016$, Hedges' g = 0.61. Therefore, H2b was supported. Together, these findings indicate that the immediate knowledge gains were not maintained at one month. Knowledge scores returned to approximately baseline levels following the intervention. The comparisons are presented in Table 2.

Table 2
Class-Level Changes in Knowledge Scores Following Educational Intervention

Comparison	***Mean Difference***	***t*(df)**	***p***	***Hedges' g***
Pre vs Post	+1.40	17.30 (17)	<.001	3.89
Pre vs Follow-up	+0.25	0.55 (17)	.589	0.12
Post vs Follow-up	−1.15	2.69 (17)	.016	0.61

Note. Paired samples t-tests were used for all comparisons. Effect sizes reported as Hedges g.

3.5 RQ4: Immediate Learning and Behavior-Change Intentions

RQ4 examined the knowledge gains and behavior-change intentions students described immediately following the educational intervention. Students described a range of intended behavior changes immediately after the intervention. The most frequently reported intention was reducing energy drink consumption (*n* = 31). Other intentions included substituting alternative beverages (*n* = 10), reducing overall caffeine intake (*n* = 9), avoiding caffeine before exercise or increasing exercise (*n* = 9), improving diet (*n* = 7), and improving sleep (*n* = 5). Some students reported no intention to change because they did not currently consume energy drinks (*n* = 47), did not plan to change (*n* = 36), or had already changed their behavior before the intervention (*n* = 11). These findings are summarized in Table 3. In response to the question asking whether they learned something new, 141 responses indicated general learning. Because responses could contain more than one idea, individual responses were coded into multiple thematic categories. The most frequently reported learning themes were safe caffeine limits (*n* = 38), cardiovascular and other health risks (*n* = 32), awareness of energy drink ingredients (*n* = 16), and safer alternatives for maintaining energy (*n* = 13). Less frequent themes included energy drink intoxication, duration of stimulant effects, anxiety-related effects, and risks associated with consuming caffeine before exercise. Seven responses indicated that no new information was learned. These findings are summarized in Table 4.

Table 3
Qualitative Themes from Immediate Post-Intervention Behavior-Change Intentions

Theme / Category	Description	Total Frequency (*n*)
No energy drink use	Do not consume energy drinks	47
No planned change	No intention to change behavior	36
Reduce energy drink intake	Decrease frequency or amount	31
Already changed behavior	Changes made prior to intervention	11
Substitute alternatives	Tea, coffee, or other beverages	10
Reduce overall caffeine	Beyond just energy drinks	9
Avoid caffeine before exercise / increase exercise	Reduce stimulant use around workouts and improve fitness	9
Improve diet	Nutrition-related changes	7
Improve sleep	Sleep-related behavior changes	5
General "Yes" (plan to change)	Non-specific intention to change behavior	13

Theme / Category	**Description**	**Total Frequency (*n*)**
General health improvements	Non-specific healthier lifestyle goals	3
Inform others	Plan to share information with others	3
Maybe / unsure	Uncertain about behavior change	1
Read nutrition labels	Increased ingredient awareness	1
Seeking medical advice	For fatigue or energy concerns	1
└ Supported decision not to use	Intervention reinforced non-use (subset)	3

Note. Responses to the question: "Do you plan on changing your behavior?" Frequencies represent thematic occurrences. Responses may be coded into multiple categories.

Table 4
Qualitative Themes from Immediate Post-Intervention Learning Responses

Theme / Category	**Description**	**Total Frequency (*n*)**
General "Yes" (learned something new)	Students reported learning something new (non-specific)	141
Safe caffeine limits	Awareness of 400 mg daily recommendation	38
Cardiovascular and health risks	Heart rate, blood pressure, arrhythmias, negative effects	32
Ingredient awareness	Guarana, taurine, hidden caffeine	16
Safer alternatives	Tea, hydration, healthier energy strategies	13
Energy drink intoxication	Severe outcomes (e.g., seizures)	4
Duration of effects	4–6-hour stimulant awareness	2
Amount of caffeine in energy drinks	Specific caffeine content knowledge	2
Anxiety effects	Links to anxiety or mental health	2
Caffeine before workout	Awareness of exercise risks	2

Theme / Category	Description	Total Frequency (*n*)
Shared information with others	Intent to tell others	1
Put things into perspective	Changed understanding/perception	1
Nothing / No learning	Reported no new information	7

Note. Responses to the question: "Did you learn something new from the educational session?" Frequencies represent the number of times a theme was identified; responses may be coded into multiple categories.

3.6 RQ5, H3a, and H3b: Changes in Energy Drink Consumption

RQ5 examined whether class-level energy drink consumption frequency and amount consumed per occasion changed from baseline to the one-month follow-up. H3a predicted that consumption frequency would decrease, while H3b predicted that the amount consumed per occasion would decrease. Mean energy drink consumption frequency decreased from 2.47 (*SD* = 0.58) at baseline to 1.51 (*SD* = 0.62) at follow-up. The mean decrease of 0.96 was statistically significant, $p = .045$. Therefore, H3a was supported. The mean amount of energy drinks consumed per occasion decreased from 1.62 (*SD* = 0.44) at baseline to 1.31 (*SD* = 0.39) at follow-up. Although the mean decreased by 0.31, the change was not statistically significant, $p = .120$. Therefore, H3b was not supported. These results indicate that students consumed energy drinks less frequently after the intervention, but the amount consumed on each occasion did not decrease significantly. The results are presented in Table 5.

Table 5
Exploratory Class-Level Changes in Health Behaviors at One-Month Follow-Up

Behavior	Pre-Mean (*SD*)	Follow-up Mean (*SD*)	Mean Difference	*p*
Energy drink frequency	2.47 (0.58)	1.51 (0.62)	−0.96	.045
Energy drink amount	1.62 (0.44)	1.31 (0.39)	−0.31	.120
Sleep duration	2.14 (0.49)	2.33 (0.52)	+0.19	.21

Note. Paired samples t-tests were used for pre to follow up comparisons.

3.7 Changes in Other Health Behaviors

Changes in broader health behaviors, including sleep duration, sleep quality, physical activity, and dietary intake, were examined from baseline to the one-month follow-up. Small directional improvements were observed across several outcomes; however, these changes were not statistically significant. For example, mean sleep duration increased from 2.14 (*SD* = 0.49) at baseline to 2.33 (*SD* = 0.52) at follow-up, but the change was not statistically significant, $p = .21$. Similar nonsignificant trends were observed for sleep quality, physical activity, and dietary behaviors. These findings suggest that the intervention had a clearer effect on energy drink consumption frequency than on broader health behaviors. More intensive interventions, repeated education, or a longer follow-up period may be needed to produce measurable changes in sleep, diet, and physical activity.

3.8 Summary of Hypothesis Testing

H1 was supported because class-level knowledge scores increased significantly immediately after the intervention. H2a was not supported because follow-up knowledge scores were not significantly higher than baseline scores. H2b was supported because knowledge scores decreased significantly from post-intervention to follow-up. H3a was supported because energy drink consumption frequency decreased significantly from baseline to follow-up. H3b was not supported because the reduction in the amount consumed per occasion was not statistically significant.

3.9 RQ6: Health-Related Behavior Changes and Harm-Reduction Strategies

Open-ended responses to the one-month follow-up question, “How has your lifestyle changed in the past month to increase health and fitness?” reflected a range of reported health-related changes. Because responses could contain more than one idea, individual responses were coded into multiple thematic categories, and frequencies represent thematic occurrences rather than unique participants.

The most frequently reported change was increased physical activity (*n* = 27), followed by reduced energy drink consumption (*n* = 23). Other commonly reported changes included improved diet (*n* = 15), improved sleep (*n* = 13), and reduced overall caffeine intake (*n* = 7). Students also reported substituting energy drinks with alternative beverages (*n* = 4), paying greater attention to nutrition labels and caffeine content (*n* = 4), and experiencing reduced stress or anxiety (*n* = 4). Five responses reflected a general positive health change without additional detail, while three indicated an intention to make changes in the future.

Not all responses reflected behavior change. Eighteen thematic occurrences indicated no change, and nine indicated that the respondent did not consume energy drinks before the intervention. Less frequent responses included increased tiredness (*n* = 1) and increased energy (*n* = 1). Overall, the qualitative findings suggest that the intervention was followed by changes extending beyond energy drink consumption, particularly in physical activity, diet, sleep, and caffeine-related decision-making. These findings are summarized in Table 6.

Table 6
Qualitative Themes from One-Month Follow-Up Lifestyle Changes

Theme / Category	Description	Total Frequency (*n*)
Increased physical activity	More exercise or movement	27
Reduced energy drink consumption	Lower frequency or stopped use	23
No change	No lifestyle differences reported	18
Improved diet	More balanced or healthier eating habits	15

Theme / Category	Description	Total Frequency (*n*)
Improved sleep	Better sleep duration or quality	13
Reduced overall caffeine	Decreased caffeine beyond energy drinks	7
No prior energy drink use	Not applicable to behavior change	9
General positive change	Non-specific improvement in health	5
Substitute alternatives	Replacing energy drinks with other options	4
Want to change in the future	Expressed intent but no current change	3
Increased label awareness	Checking ingredients or caffeine content	4
Reduced stress/anxiety	Mental health improvements	4
Increased tiredness	Reported feeling more tired	1
Increased energy	Reported improved energy levels	1

Note. Responses to the question: "How has your lifestyle changed in the past month to increase health and fitness?" Frequencies represent thematic occurrences. Responses may be coded into multiple categories.

Discussion

4.1 Interpretation of Results

The intervention produced a significant immediate increase in class-level knowledge, supporting H1. This finding is consistent with literature showing that consumers may not fully understand the caffeine content, stimulant combinations, and cardiovascular risks associated with energy drinks (Higgins et al., 2018; NCCIH, 2023). The present study extends this literature by showing that a brief, low-resource classroom intervention can substantially improve students' immediate awareness of caffeine limits, ingredients, and health risks. However, knowledge scores declined significantly from post-intervention to the one-month follow-up and were not significantly higher than baseline. Therefore, H2a was not supported, while H2b was supported. This suggests that a single session may improve immediate learning but may not be sufficient to maintain detailed knowledge over time. Reinforcement through booster sessions or follow-up materials may be needed to support longer-term retention.

Despite the decline in knowledge, energy drink consumption frequency decreased significantly at follow-up, supporting H3a. This finding is important because frequent energy drink use may contribute to increased heart rate, blood pressure, palpitations, and other cardiovascular concerns (Higgins et al., 2018; Martínez-Sánchez et al., 2023; UC Davis Health, 2024). The results extend prior research by suggesting that brief education may influence consumption behavior even when detailed knowledge is not fully retained. In contrast, the amount consumed per occasion did not decrease significantly, so H3b was not supported. The intervention may therefore have influenced how often students consumed energy drinks more strongly than how much they consumed at one time. The findings are also consistent with the Health Belief Model. Information about cardiovascular risks was intended to increase perceived susceptibility and severity, while healthier alternatives, caffeine-label reading, sleep, hydration, and nutrition strategies emphasized the benefits of change and reduced practical barriers (Alyafei & Easton Carr, 2024). The educational session also served as a cue to action. The reduction in consumption frequency suggests that these components may have encouraged behavior change even though detailed knowledge declined over time.

4.2 Public Health Relevance

Energy drink consumption is common among college students and is often used to manage fatigue, academic pressure, and sleep deprivation. Because excessive stimulant intake can increase heart rate and blood pressure, even modest reductions in use may support cardiovascular risk reduction among young adults. This study suggests that a brief classroom education session can reduce how often students consume energy drinks and encourage healthier choices, such as reading caffeine labels, choosing alternatives, and prioritizing sleep and nutrition. These findings support the use of short, evidence-based education in campus wellness programs, general education courses, health courses, or first year experience programs.

4.3 Limitations of the Study

Several limitations should be considered when interpreting these findings. Analyses were conducted at the class level because anonymous responses could not be linked across assessment periods. As a result, the findings describe average changes across participating classes and cannot determine whether the same individual students experienced changes in knowledge or behavior. Aggregation may also have reduced individual variability and obscured differences between students who responded strongly to the intervention and those who experienced little or no change. Accordingly, the results should be interpreted as class-level intervention effects rather than individual-level changes.

The relatively small number of participating classes reduced statistical power and limited the range of analyses that could be performed. All behaviors were self-reported and may have been affected by recall or social desirability bias. The one-month follow-up period also limited the ability to assess whether behavior changes were sustained over a longer period. In addition, no physiological measures, such as blood pressure or heart rate, were collected; therefore, health effects were inferred from reported behavior changes rather than measured directly. Because the intervention consisted of a single session without reinforcement, the decline in knowledge may reflect limited continued exposure rather than ineffective educational content. Finally, the study was conducted at one university and included students from health-related courses, which may limit generalizability and introduce selection bias related to interest in health topics.

4.4 Future Research Directions

Future studies should include a larger number of classes to increase statistical power and allow comparison across course types or academic disciplines. Repeating or reinforcing the educational content through booster sessions, reminders, or supplemental materials may improve long term knowledge retention while maintaining or enhancing behavior change. Longer follow-up periods are needed to determine whether reductions in energy drink use, and related lifestyle improvements persist over time. Future work could also incorporate objective or clinical measures of health, such as blood pressure or sleep tracking, to better assess physiological effects of reduced stimulant use. With larger aggregated samples, additional class level analyses such as comparisons across departments or repeated measures models over multiple follow-up points could be conducted. Although this study intentionally focused on class-level data, complementary research in other settings could examine individual-level change trajectories to better understand variability in response to the intervention.

Conclusion

This study found that a brief classroom-based educational intervention improved students' knowledge of energy drink risks immediately after the session and was associated with reduced energy drink consumption frequency at one month. Although knowledge gains declined over time, students still reported behavior changes such as reducing energy drink use, choosing alternatives, reading caffeine labels, and improving sleep or nutrition habits. These findings suggest that simple risk communication and realistic harm reduction strategies may influence student decision making even when detailed factual recall declines. This study contributes to health education research by showing that a brief, low resource intervention can be implemented during regular class time and may support healthier energy drink behaviors among college students. Institutions of higher education should consider integrating energy drink and caffeine safety education into existing campus wellness efforts, health courses, or first year student programming. Future programs may benefit from booster sessions or reminders to improve knowledge retention and support long term behavior change. Expanding this intervention across additional courses and student populations could help strengthen evidence for scalable approaches to reducing risky stimulant use and promoting cardiovascular health.

References

Ali, F., Rehman, H., Babayan, Z., Stapleton, D., & Joshi, D. D. (2022). Energy drinks and their impact on the cardiovascular system: Potential mechanisms. *Frontiers in Public Health, 10*, 959392. https://pmc.ncbi.nlm.nih.gov/articles/PMC9593923/

Alyafei, A., & Easton Carr, R. (2024). The health belief model of behavior change. In *StatPearls*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK606120/

Center for Science in the Public Interest. (2024). *Caffeine chart*. https://www.cspi.org/caffeine-chart#energy_drink

Centers for Disease Control and Prevention. (2024). *Leading causes of death*. U.S. Department of Health and Human Services. https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

Higgins, J. P., Babu, K., Deuster, P. A., & Shearer, J. (2018). Energy beverages: Content and safety. *Mayo Clinic Proceedings, 93*(12), 1694–1706. https://pmc.ncbi.nlm.nih.gov/articles/PMC5714807/

Jefferson Health. (2023). *Can energy drinks hurt your heart?* https://www.jeffersonhealth.org/your-health/living-well/can-energy-drinks-hurt-your-heart

Martínez-Sánchez, J. M., Martín-Aragón, S., & Sánchez-Muniz, F. J. (2023). Effects of energy drink consumption on cardiovascular health: A review. *Journal of Food Composition and Analysis, 122*, 105533. https://www.sciencedirect.com/science/article/abs/pii/S2352007823003153

Mayo Clinic Staff. (2024). *Caffeine: How much is too much?* Mayo Foundation for Medical Education and Research. https://www.mayoclinic.org/healthy-lifestyle/nutrition-and-healthy-eating/in-depth/caffeine/art-20045678

National Center for Complementary and Integrative Health. (2023). *Energy drinks*. National Institutes of Health. https://www.nccih.nih.gov/health/energy-drinks

UC Davis Health. (2024, May). *How do energy drinks affect your heart and overall health?* University of California. https://health.ucdavis.edu/blog/good-food/how-do-energy-drinks-affect-your-heart-and-overall-health/2024/05

7. Appendices

7.1 Survey instruments

The pre-intervention, post-intervention, and follow-up survey instruments are presented in Table 7.

Table 7
Pre/Post/Follow-Up Surveys

Pre-/Post/ Follow-Up	Question	Answer Options
Pre/Follow-Up	Age	Free-Response
Pre/Follow-Up	Sex	Free-Response
Pre/Follow-Up	Classification	Free-Response
Pre/Follow-Up	Major/Minor	Free-Response
Pre/Follow-Up	In a typical week, on how many days do you do aerobic activity (e.g., brisk walking, running, sports) for ≥10 minutes at a time?	a. 0 b. 1-2 c. 3-4 d. 5-6 e. 7
Pre/Follow-Up	On those days, about how many total minutes do you accumulate?	a. 0–9 b. 10–29 c. 30–59 d. 60–89 e. 90–149 f. 150–299 g. 300+
Pre/Follow-Up	In a typical week, on how many days do you do muscle-strengthening activities (weights, resistance bands, bodyweight, etc.)?	a. 0 b. 1-2 c. 3-4 d. 5-6 e. 7
Pre/Follow-Up	On a typical day, how many servings of fruits and vegetables do you eat? (1 serving ≈ 1 cup raw veg or fruit; ½ cup cooked veg)	a. 0 b. 1 c. 2 d. 3 e. 4+
Pre/Follow-Up	In a typical week, how many meals/snacks do you get from fast food or fried takeout?	a. 0 b. 1-2 c. 3-4 d. 5-6 e. 7+
Pre/Follow-Up	On a typical night, how many hours of sleep do you get?	a. Less than 5 b. 5-6 c. 7-8 d. 9-10 e. 11+

Pre-/Post/ Follow-Up	Question	Answer Options
Pre/Follow-Up	Over the past 2 weeks, how would you rate your sleep quality?	a. Very poor b. Poor c. Fair d. Good e. Very Good
Pre/Follow-Up	In the past 30 days, how often did you consume energy drinks?	a. Never b. 1-2 days c. 1 day/week d. 2-3 days/week e. 4-6 days/week f. Daily
Pre/Follow-Up	On the days you consume energy drinks, how many cans/bottles do you usually have?	a. None b. ½ or less c. 1 d. 2 e. 3 f. 4+
Pre-/Post/ Follow-Up	What is considered a safe maximum daily intake of caffeine for most adults?	a. 100 mg b. 200 mg c. 300 mg d. 400 mg Answer: D
Pre-/Post/ Follow-Up	What cardiovascular effect is linked to high caffeine and stimulant levels in energy drinks?	a. Decreased heart rate and blood pressure. b. Increased heart rate and blood pressure c. Stronger bones and muscles d. Lower cholesterol levels Answer: B
Pre-/Post/ Follow-Up	What is a healthier alternative to energy drinks before working out?	a. Coffee with extra sugar b. Black or green tea, or electrolyte drinks, in moderation c. Energy drinks with taurine only d. Doubling up on caffeine for faster results Answer: B
Pre-/Post/ Follow-Up	Ingredients like guarana add additional caffeine to energy drinks.	a. True b. False c. Not sure Answer: A

Pre-/Post/ Follow-Up	**Question**	**Answer Options**
Pre-/Post/ Follow-Up	Excess caffeine from energy drinks is associated with which condition?	a. Stronger heart contractions b. Palpitations and arrhythmias such as A-Fib c. Improved oxygen delivery to tissues d. Lower stroke risk Answer: B
Pre-/Post/ Follow-Up	How can high sugar levels in energy drinks indirectly affect cardiovascular health?	a. By improving hydration b. By contributing to obesity, a major risk factor for heart disease c. By lowering cholesterol d. By increasing lung capacity Answer: B
Pre-/Post/ Follow-Up	What did research suggest about artificial sweeteners in energy drinks?	a. They have no cardiovascular effects. b. Higher consumption is associated with increased risk of cardiovascular and cerebrovascular diseases. c. They protect against arrhythmias. d. They strengthen heart muscles. Answer: B
Pre-/Post/ Follow-Up	Why is it risky to consume energy drinks before exercise?	a. Exercise and stimulants together can exaggerate the strain on the heart. b. They improve endurance too much. c. They make blood pressure drop suddenly. d. They reduce the heart rate below safe levels. Answer: A
Pre-/Post/ Follow-Up	Severe intoxication from energy drinks has been linked to which medical events?	a. Arthritis and joint pain b. Seizures and arrhythmias c. Osteoporosis d. Hair loss Answer: B

Pre-/Post/ Follow-Up	Question	Answer Options
Pre-/Post/ Follow-Up	Frequent consumption of energy drinks may lead to:	a. Long-term cardiovascular stress and dependence b. Improved sleep quality c. Permanent elimination of hypertension d. Lower risk of strokes Answer: A
Follow-up	In the past month, did you change your energy drink habits because of the program?	a. No change b. Reduced frequency c. Reduced amount per occasion d. Stopped entirely e. Other (please explain) No current answer
Post	Did you learn something new?	Free-Response
Post	Do you plan on changing your behavior?	Free-Response
Follow-up	How has your lifestyle changed in the past month to increase health and fitness?	Free-Response

Socioeconomic and Demographic Predictors of Asthma Exacerbations Among U.S. Adults: Analysis of NHANES 2021-2023

Rachel Barker
Department of Public Health
Sam Houston State University

Kiran Sapkota
Department of Public Health
Sam Houston State University

Abstract

Asthma remains a significant public health concern among adults in the United States and continues to contribute to substantial morbidity despite advances in treatment and disease management. This study examined demographic, socioeconomic, and health-related predictors of asthma attacks among adults with current asthma.

A cross-sectional analysis was conducted using 2021-2023 National Health and Nutrition Examination Survey data from 250 adults aged 18 years and older with current asthma. Weighted descriptive statistics, chi-square tests, independent-samples t tests, and complex-samples logistic regression were used to assess associations between asthma attacks in the past 12 months and demographic, socioeconomic, behavioral, clinical, and insurance-related factors.

Nearly half of adults with asthma (47.3%) reported an asthma attack in the past year. Female sex was the strongest predictor of asthma exacerbation, with women demonstrating significantly higher odds of experiencing an asthma attack than men (adjusted odds ratio = 4.11, 95% confidence interval = 1.63-10.35). Lower educational attainment and a lower poverty-to-income ratio were also significantly associated with higher odds of asthma attacks. Health insurance coverage was not significantly associated with asthma attacks after adjustment for covariates. Body mass index, smoking, age, diabetes, hypertension, and hay fever were not statistically significant predictors in the adjusted model.

These findings highlight the importance of addressing social determinants of health and socioeconomic disparities in asthma prevention and management efforts.

Key Words: asthma; social determinants of health; NHANES; adult exacerbations; adult asthma

Introduction

1.1 Background and Epidemiology of Asthma

Asthma is a chronic respiratory disease that affects millions of adults in the United States and worldwide, contributing substantially to morbidity, reduced quality of life, and healthcare utilization (Centers for Disease Control and Prevention [CDC], 2023). Characterized by chronic airway inflammation, bronchoconstriction, and episodic exacerbations, asthma can interfere with daily functioning, occupational performance, sleep quality, and physical activity (Global Initiative for Asthma

[GINA], 2022). Although advances in pharmacologic treatment and disease management have improved symptom control for many individuals, asthma exacerbations remain common among adults and continue to contribute to emergency department visits, hospitalizations, and preventable healthcare costs (Syamlal et al., 2022).

Asthma also represents an important public health concern because its burden is not distributed equally across populations. Prior research has demonstrated substantial demographic and socioeconomic disparities in asthma prevalence, disease severity, and exacerbation risk (Akinbami et al., 2019). Women consistently experience higher asthma morbidity and are more likely to report severe symptoms and asthma-related healthcare utilization compared with men (GINA, 2022). Similarly, individuals with lower-income and lower educational attainment often experience worse asthma outcomes due to increased exposure to environmental triggers, reduced access to healthcare resources, lower health literacy, and barriers to effective disease management (Hagan et al., 2021).

Beyond direct medical costs, asthma produces substantial indirect economic consequences through missed workdays, reduced productivity, and diminished overall quality of life. Adults with poorly controlled asthma frequently report activity limitations, disrupted sleep, and psychological stress associated with unpredictable exacerbations (CDC, 2023). These patterns highlight the importance of examining not only clinical risk factors, but also the broader demographic and socioeconomic conditions that may influence asthma control and exacerbation risk among U.S. adults.

1.2 Gaps in Existing Research

Although substantial research has examined asthma prevalence, environmental triggers, and pediatric asthma management, fewer studies have focused specifically on adult asthma exacerbations using recent nationally representative datasets. Much of the existing literature emphasizes childhood asthma, regional populations, or clinical samples, which may limit the generalizability of findings to the broader U.S. adult population (Akinbami et al., 2019). In addition, many previous studies focus primarily on asthma prevalence or healthcare access rather than acute exacerbation outcomes such as asthma attacks, emergency department visits, or loss of disease control.

Existing research has also frequently examined healthcare access and insurance coverage independently, with less attention given to the combined influence of demographic and socioeconomic factors such as sex, educational attainment, and income level on adult asthma outcomes (Ward et al., 2021). However, growing evidence suggests that broader social determinants of health may substantially influence asthma management and exacerbation risk through mechanisms including environmental exposures, reduced health literacy, occupational risks, and financial barriers to preventive care (Hagan et al., 2021).

Additionally, relatively few studies have utilized recent NHANES cycles reflecting contemporary healthcare systems and socioeconomic conditions following major healthcare policy changes and insurance expansions in the United States. Updated analyses using nationally representative data are needed to better understand the demographic, socioeconomic, and healthcare-related predictors associated with asthma exacerbations among U.S. adults.

Addressing these gaps may provide clearer insight into the structural and social factors influencing asthma control and may help inform targeted public health interventions aimed at reducing disparities in asthma morbidity and healthcare utilization.

1.3 Study Purpose and Research Question

The primary objective of this study was to examine demographic, socioeconomic, behavioral, clinical, and healthcare-related predictors of asthma exacerbations among U.S. adults with current asthma using data from the 2021 to 2023 National Health and Nutrition Examination Survey (NHANES). Particular attention was given to the influence of sex, educational attainment, poverty-to-income ratio, and health insurance coverage on the likelihood of experiencing an asthma attack within the past 12 months.

This study utilized nationally representative NHANES questionnaire and examination data to evaluate associations between asthma attacks and multiple demographic and health-related variables among adults with current asthma. Variables included in the analysis consisted of sex, race and ethnicity, age, educational attainment, poverty-to-income ratio, smoking status, body mass index, hypertension, diabetes, hay fever, and current health insurance coverage.

The research question guiding this study was: Which demographic, socioeconomic, behavioral, clinical, and healthcare-related factors are associated with experiencing an asthma attack in the past 12 months among U.S. adults with current asthma?

By examining these relationships using recent nationally representative data, this study seeks to contribute to the growing body of literature addressing social determinants of health and disparities in asthma outcomes among adults. Understanding the factors associated with asthma exacerbations may help inform targeted public health interventions and improve long-term asthma-management strategies for vulnerable populations.

Literature Review

2.1 Asthma Management, Healthcare Access, and Social Determinants

Effective asthma management depends on consistent access to healthcare, preventive services, medications, and patient education. Individuals with well-controlled asthma are more likely to receive routine monitoring, appropriate medication adjustments, and asthma action plans that support symptom management and reduce exacerbation risk (Global Initiative for Asthma [GINA], 2022). Preventive care and continuity of care are especially important for chronic respiratory diseases because inadequate disease management may contribute to worsening symptoms, emergency department visits, hospitalizations, and reduced quality of life (CDC, 2023).

Although healthcare access remains an important component of asthma management, growing evidence suggests that broader social determinants of health may substantially influence asthma outcomes among adults (Hagan et al., 2021). Social determinants of health refer to the conditions in which individuals are born, live, work, learn, and access healthcare, all of which can affect disease risk and long-term health outcomes (World Health Organization [WHO], 2023). Factors such as educational attainment, income level, housing quality, occupational exposures, transportation access, and health literacy may shape an individual's ability to effectively manage asthma and avoid exacerbations.

Educational attainment may influence asthma outcomes through health literacy, medication adherence, and understanding of disease management strategies. Adults with lower educational attainment may experience greater difficulty interpreting asthma action plans, recognizing worsening symptoms, or navigating complex healthcare systems (Smith et al., 2021). Similarly, lower-income populations may face barriers including financial strain, unstable housing conditions, environmental pollutant exposure, and limited access to preventive healthcare resources, all of which may contribute to poor asthma control and increased exacerbation risk (Liu et al., 2021).

These patterns suggest that asthma outcomes are influenced not only by biological or clinical factors, but also by broader structural and socioeconomic conditions. Examining demographic and socioeconomic predictors alongside healthcare access variables may therefore provide a more comprehensive understanding of factors associated with asthma exacerbations among U.S. adults.

2.2 Social Determinants and Asthma Disparities

Substantial evidence demonstrates that asthma outcomes are strongly influenced by social determinants of health and broader socioeconomic conditions. Individuals with lower-income levels are more likely to experience environmental and structural exposures that may worsen asthma symptoms and increase exacerbation risk. Poor housing quality, mold exposure, pest infestations, air pollution, and proximity to high-traffic areas have all been associated with increased asthma morbidity among adults (Liu et al., 2021). Financial barriers may also limit access to medications, transportation, preventive healthcare visits, and other resources necessary for effective long-term asthma management.

Educational attainment is another important predictor of asthma outcomes. Lower levels of education are frequently associated with reduced health literacy, lower medication adherence, and decreased understanding of asthma self-management practices (Smith et al., 2021). Adults with lower educational attainment may experience greater difficulty recognizing worsening symptoms, following asthma action plans, or navigating healthcare systems effectively. These challenges may contribute to delayed treatment-seeking behaviors and increased risk of asthma exacerbations.

Sex based disparities in asthma outcomes have also been consistently documented within the literature. Although asthma is more common among boys during childhood, adult women experience higher asthma prevalence, greater symptom severity, and more frequent exacerbations compared with men (Akinbami et al., 2019). Women are also more likely to report asthma-related healthcare utilization, including emergency department visits and hospitalizations (Global Initiative for Asthma [GINA], 2022). Researchers have proposed several contributing mechanisms, including hormonal influences, immune response differences, psychosocial stressors, and environmental exposures (Postma, 2007).

Occupational and workplace exposures may further contribute to asthma disparities among adults. Individuals employed in occupations involving dust, cleaning chemicals, industrial fumes, and other respiratory irritants may face increased risk of airway inflammation and asthma exacerbations (Henneberger et al., 2015). These occupational risks may disproportionately affect lower-income populations and individuals employed in manual labor or service industry positions with limited workplace protections.

Collectively, these findings demonstrate that asthma control is shaped not only by clinical treatment, but also by the broader social, environmental, and structural conditions that influence daily living circumstances and healthcare access. Understanding these disparities is essential for developing public health interventions aimed at reducing asthma morbidity and improving long-term disease-management outcomes.

2.3 Healthcare Access and Chronic Disease Outcomes

Healthcare access remains an important component of chronic disease prevention and management. Individuals with consistent access to healthcare services are more likely to receive preventive care, maintain continuity with healthcare providers, obtain prescribed medications, and participate in routine disease monitoring (Schoen et al., 2017). For chronic respiratory diseases such as asthma, ongoing management is especially important because inadequate treatment and delayed care may contribute to worsening symptoms, increased healthcare utilization, and greater risk of exacerbation (Global Initiative for Asthma [GINA], 2022).

Health insurance coverage may influence asthma management by improving access to primary care providers, specialist services, prescription medications, and preventive treatment resources (Tang et al., 2020). Conversely, uninsured individuals may delay medical visits, experience interruptions in medication access, or rely more heavily on emergency departments for episodic treatment rather than long-term disease-management (Ward et al., 2021). However, healthcare access alone may not fully explain disparities in asthma outcomes, as socioeconomic and structural factors frequently interact with healthcare availability to shape disease control and exacerbation risk.

Emerging public health research increasingly emphasizes that chronic disease outcomes are influenced by a combination of healthcare access, demographic characteristics, socioeconomic conditions, and environmental exposures (Hagan et al., 2021). Even among insured populations, disparities in education, income, housing quality, occupational exposure, and health literacy may contribute to differences in asthma management and overall respiratory health. These findings suggest that effective asthma prevention and management strategies require approaches that address both healthcare access and the broader social determinants that influence long-term health outcomes.

Methods

3.1 Study Design and Data Source

This cross-sectional study used data from the 2021-2023 continuous National Health and Nutrition Examination Survey (NHANES). NHANES is conducted by the Centers for Disease Control and Prevention's National Center for Health Statistics and uses a complex, multistage probability sampling design to produce nationally representative estimates of the civilian, noninstitutionalized U.S. population. Data are collected through interviews, physical examinations, and laboratory assessments.

The following NHANES datasets were merged for this analysis: DEMO_L (Demographics), MCQ_L (Medical Conditions), HIQ_L (Health Insurance), BMX_L (Body Measures), SMQ_L (Smoking), BPQ_L (Hypertension), and DIQ_L (Diabetes).

NHANES is widely regarded as a gold standard dataset for population health research due to its rigorous sampling framework and comprehensive data collection procedures. The survey intentionally oversamples certain demographic groups, including racial and ethnic minorities and older adults, to improve estimate precision for populations that may otherwise be underrepresented. This design strengthens the reliability of subgroup analyses and enhances the external validity of findings. Additionally, the combination of interview, examination, and laboratory components allows researchers to evaluate health outcomes using both self-reported and objective indicators, making NHANES particularly valuable for chronic disease research and health disparities investigations.

3.2 Study Population

The analytic sample included adults aged 18 years and older who participated in NHANES between 2021 and 2023. Participants were included if they reported a current diagnosis of asthma, defined as answering “Yes” to both MCQ010 (“Has a doctor or other health professional ever told you that you have asthma?”) and MCQ020 (“Do you still have asthma?”). Individuals were excluded if they were younger than 18 or had missing or invalid responses for key variables. The final analytic sample consisted of 250 adults with current asthma, representing an estimated 7.4 million adults in the United States.

Focusing specifically on adults with current asthma allowed this study to examine exacerbation risk within a clinically relevant subgroup rather than the general population. Restricting the sample to individuals who both reported a professional diagnosis and indicated ongoing asthma status improved internal validity by reducing misclassification bias. Although the final analytic sample size was modest, the application of NHANES sampling weights ensured that estimates reflected millions of adults nationally. This targeted approach provided a clearer understanding of factors influencing disease control among individuals actively managing asthma rather than those with only historical or childhood diagnoses.

3.3 Variables and Measures

The outcome variable was an asthma attack in the past 12 months (MCQ040), defined as answering “Yes” to the question, “During the past 12 months, have you had an episode of asthma or an asthma attack?”

The variable HIQ210, which captures any time without insurance in the past year, showed extremely low variability in the merged NHANES cycles and resulted in quasi-complete separation during preliminary analysis. Because this prevented SPSS from estimating logistic regression coefficients in the complex-samples model, HIQ210 was removed from the analytic model. Current health insurance coverage at the time of the interview (HIQ011) was retained as the main insurance variable.

Based on the guidance provided during statistical review and data availability, predictors with stronger epidemiologic relevance and sufficient variation were included in the final analysis. The full list of predictor variables and corresponding NHANES codes is provided in Appendix Table 1.

3.4 Statistical Analysis

Prior to multivariable modeling, all predictor variables were examined for variability, distribution, and suitability for inclusion in regression analyses. This preliminary screening step ensured that variables met statistical assumptions and contained sufficient variation for reliable estimation. During model building, smoking history (SMQ020) showed no variability among adults in the final analytic sample, with all respondents reporting ever smoking at least 100 cigarettes, and therefore could not be estimated in the multivariable model. SMQ020 was retained for descriptive analyses but excluded from the final regression model.

Descriptive statistics were used to summarize demographic characteristics, asthma outcomes, and predictor variables. Means and standard deviations were calculated for continuous variables, including age, body mass index (BMI), and poverty-to-income ratio. Weighted frequencies and percentages were reported for categorical variables to reflect the complex survey design.

Bivariate analyses were conducted to examine unadjusted associations between an asthma attack in the past 12 months and each predictor. Independent-samples t tests were used for continuous variables, and chi-square tests were used for categorical variables.

A multivariable complex-samples logistic regression model was used to examine factors associated with experiencing an asthma attack in the past 12 months among adults with current asthma. The model excluded HIQ210 due to quasi-complete separation and included alternative predictors with adequate variability. Adjusted odds ratios and 95% confidence intervals were obtained for each predictor. Because the Complex Samples procedure in SPSS treats the highest value of the dependent variable as the reference category, model coefficients were interpreted so that reported odds ratios represent the odds of having an asthma attack rather than the odds of no attack.

All analyses incorporated NHANES sample weights, strata, and primary sampling units to account for the complex sampling design. The full-sample 2-year interview weight (WTINT2YR) was used as the analysis weight, with SDMVSTRA as the strata variable and SDMVPSU as the primary sampling unit variable. Analyses were conducted using SPSS Complex Samples with a two-sided significance level of $p < .05$.

Logistic regression was selected as the primary analytical technique because the outcome variable, asthma attack in the past 12 months, was dichotomous in nature and required estimation of the probability of occurrence based on multiple predictors (Hosmer et al., 2013). This method allows for simultaneous adjustment of demographic, socioeconomic, behavioral, and clinical variables, providing a more comprehensive understanding of independent associations than bivariate analysis alone. The use of a multivariable approach also reduces the likelihood that observed relationships are confounded by overlapping predictors such as income, education, or age.

Incorporating the NHANES complex sampling design into statistical analyses is essential for producing unbiased national estimates. NHANES employs stratification, clustering, and oversampling of specific demographic groups to ensure adequate representation of minority populations. Failure to apply sampling weights and design variables can result in distorted parameter estimates and inaccurate standard errors. This study utilized the SPSS Complex Samples module in SPSS and applied interview weights, strata, and primary sampling unit variables. This approach preserved the integrity of the national sampling framework and improved the generalizability of findings to the broader U.S. adult population.

While cross-sectional data provide valuable insight into associations between variables, they do not establish causation or temporal sequencing. Consequently, the relationships identified in this study should be interpreted as correlational rather than causal. Nonetheless, cross-sectional national surveys remain a widely accepted and efficient approach for examining population health trends and generating hypotheses for longitudinal investigation.

3.5 Ethical Considerations

NHANES data are publicly available, de-identified, and collected under the approval of the National Center for Health Statistics Research Ethics Review Board. Because this study involved secondary analysis of existing data, additional Institutional Review Board approval was not required.

All NHANES participants provide informed consent prior to participation, and data are released only after extensive de-identification procedures that remove personally identifiable information. Researchers accessing the dataset agree to ethical use guidelines that prohibit attempts to re-identify participants. Because the analysis relied exclusively on publicly available secondary data with no direct interaction with human subjects, the study posed minimal risk. Nonetheless, ethical research practice required careful handling of sensitive health information and accurate representation of findings to avoid stigmatization of vulnerable populations disproportionately affected by respiratory disease.

Results

4.1 Sample Characteristics

The final analytic sample included 250 adults with current asthma, representing approximately 7.4 million U.S. adults. The weighted mean age was 53.0 years, the mean BMI was 32.1 kg/m^2, and the mean poverty-to-income ratio was 2.56 (Table 1).

Overall, 47.3% of adults with asthma reported at least one asthma attack in the past 12 months. Most respondents reported current health insurance coverage (90.6%; Table 1). All respondents reported a lifetime smoking history of at least 100 cigarettes, and 39.0% were current smokers. About one-third (34.9%) reported hay fever in the past year (Table 1).

Regarding sociodemographic characteristics, 54.2% were female and 45.8% were male. Nearly two-thirds were non-Hispanic White, approximately 10% were non-Hispanic Black, and the remainder identified as Mexican American, other Hispanic, Asian, or multiracial (Table 1). A total of 14.7% had a college degree, 41.2% reported completed college credits with no college degree, 33.5% had completed high school or earned a GED, and 10.6% had less than a high school education. Additionally, 22.8% reported diabetes and 48.0% reported hypertension (Table 1).

4.2 Bivariate Associations

In chi-square analyses, several categorical predictors were significantly associated with asthma attacks in the past year. Sex was significantly associated with asthma attacks, with females more likely than males to report an attack ($p < .001$; Table 1). Race and ethnicity, smoking history, and hypertension status also showed significant bivariate associations. In contrast, diabetes and health insurance coverage were not significantly associated with asthma attacks. Independent samples t-tests indicated no significant differences in mean age or BMI by asthma attack status (Table 1).

Table 1.

Weighted Sample Characteristics by Asthma Attack Status

Characteristic	**Overall (N = 250)**	**No asthma attack**	**Asthma attack**	**p value**
Age, years	53.0	—	—	.601
Sex				**< .001**
Male	45.8%	60.3%	39.7%	—
Female	54.2%	47.5%	52.5%	.005
Race and ethnicity				**.934**
Mexican American	2.5%	32.4%	67.6%	—
Other Hispanic	6.0%	43.2%	56.8%	—
Non-Hispanic White	63.2%	53.2%	46.8%	—
Non-Hispanic Black	10.2%	56.4%	43.6%	—
Non-Hispanic Asian	1.3%	68.8%	31.3%	—
Other or multiracial	16.8%	48.3%	51.7%	—
Education level				**.239**
College degree or above	14.7%	—	—	—
Some college or associate degree	41.2%	—	—	.092
High school diploma or GED	33.5%	—	—	.110
Less than high school	10.6%	—	—	.044
Poverty-to-income ratio	2.56	—	—	.030
BMI, kg/m²	32.1	—	—	.126
Smoking status				**—**
Current smoker	39.0%	—	—	.548
Former or never smoker	61.0%	—	—	—
Hay fever in past 12 months	34.9%	—	—	.085
Diabetes	22.8%	55.6%	44.4%	.951
Hypertension	48.0%	47.8%	52.2%	.590
Currently insured	90.6%	52.3%	47.7%	.301

Characteristic	Overall (N = 250)	No asthma attack	Asthma attack	p value
Asthma attack in past 12 months	47.3%	0%	100%	—

Note. Percentages and p values are reproduced from the source manuscript. BMI = body mass index; GED = General Educational Development.

4.3 Multivariable complex-samples logistic regression

The multivariable complex-samples logistic regression model was statistically significant, and the Nagelkerke pseudo-R^2 indicated that approximately 26% of variability in asthma attacks was explained by the model (Table 3). Insurance status was not a significant predictor of asthma attacks after adjustment for covariates. Compared with insured adults, those without insurance had higher but highly imprecise adjusted odds (adjusted odds ratio ≈ 2.82, 95% CI 0.36 to 22.22, $p = .301$), and the confidence interval crossed the null (Table 2).

Several sociodemographic variables remained significant in the adjusted model. Females had higher adjusted odds of asthma attacks than males (adjusted odds ratio = 4.11, 95% CI 1.63 to 10.35, $p = .005$; Table 2). Education was also associated with asthma attacks. Adults with less than a high school education had higher odds compared with college graduates (adjusted odds ratio = 4.58, 95% CI 1.05 to 20.01, $p = .044$; Table 2). Differences among other education categories were not statistically significant.

Poverty-to-income ratio was inversely associated with asthma attacks. Each one-unit increase in the poverty-to-income ratio was associated with lower odds of an asthma attack (adjusted odds ratio = 0.67, 95% CI 0.46 to 0.96, $p = .030$; Table 2). Other clinical and behavioral variables were not statistically significant in the adjusted model. Hay fever showed a positive but non-significant association with asthma attacks (adjusted odds ratio ≈ 1.82, 95% CI 0.91 to 3.65, $p = .085$). Diabetes, hypertension, BMI, current smoking, and age were not significant predictors after adjusting for all covariates (Table 2).

Table 2.
Adjusted Odds Ratios for Predictors of Asthma Attacks

Predictor	Adjusted odds ratio	95% confidence interval	p value
Female sex	4.11	1.63-10.35	.005
Less than high school education	4.58	1.05-20.01	.044
Poverty-to-income ratio	0.67	0.46-0.96	.030
Insurance status	2.82	0.36-22.22	.301
Hay fever	1.82	0.91-3.65	.085

Table 3.
Model Fit Statistics for Logistic Regression

Statistic	Value
Nagelkerke R^2	26%

5. Discussion

The purpose of this study was to identify demographic, socioeconomic, and health-related predictors of an asthma attack in the past 12 months among U.S. adults with current asthma using 2021-2023 NHANES data. This study contributes updated nationally representative evidence on adult asthma exacerbations, an outcome examined less frequently than pediatric asthma or findings from regional samples. Nearly half of adults with current asthma reported an exacerbation in the past year, highlighting the continued public health burden of asthma despite improvements in available treatments (CDC, 2023). Improving asthma outcomes requires coordinated efforts that address both medical management and the broader social conditions influencing patient health. This discussion interprets the major findings, connects them to prior research, evaluates their implications, reviews limitations, and outlines opportunities for future investigation.

5.1 Principal Findings

Several predictors demonstrated clear associations with asthma attacks. Consistent with prior research, female sex emerged as the strongest predictor in the adjusted model. Women had significantly higher odds of a recent asthma attack compared with men, echoing established evidence that women experience greater asthma morbidity, possibly due to hormonal, physiological, or behavioral mechanisms (Akinbami et al., 2019; GINA, 2022).

Socioeconomic indicators were also important. Adults with lower educational attainment showed elevated odds of asthma attacks compared with college graduates, and higher poverty-to-income ratios were associated with reduced odds of an attack. These findings reinforce the role of social determinants and suggest that limited resources, reduced health literacy, and greater exposure to environmental risks may influence asthma control (Grossman et al., 2019).

Several other variables were not significant in the adjusted model. The nonsignificant result for insurance status likely reflects the limited variation in this sample, where most participants were insured. The restricted distribution reduced statistical power and prevented meaningful evaluation of insurance's role, despite prior evidence showing its importance in asthma management (Tang et al., 2020). Similarly, smoking, BMI, age, and chronic conditions such as diabetes and hypertension did not remain significant predictors. These nonsignificant findings suggest that, among adults already diagnosed with asthma, socioeconomic context and sex may play larger roles in shaping exacerbation risk than several traditional clinical or behavioral factors.

5.2 Comparison with Previous Research

The present findings align closely with previous studies documenting higher asthma morbidity among women (Akinbami et al., 2019; GINA, 2022). The significant associations with education and poverty mirror research by Grossman et al. (2019) and Ward et al. (2021), who identified socioeconomic disadvantage as a driver of asthma disparities across U.S. populations.

The absence of associations for smoking or BMI differs somewhat from studies linking these factors to poor asthma control. These discrepancies may reflect sample characteristics, limited statistical power, or differences in NHANES survey cycles. Likewise, the inability to detect insurance-related disparities contradicts prior research emphasizing insurance continuity as a critical factor in asthma management. This divergence appears to stem from measurement constraints rather than a true lack of effect, underscoring the need for richer insurance variables in future analyses.

5.3 Public Health Implications

The findings highlight opportunities for targeted intervention. The strong association between female sex and asthma attacks suggests the value of sex-informed asthma strategies, such as enhanced symptom monitoring during hormonal transitions, tailored patient education, and improved provider awareness of sex-based differences in asthma experience. The influence of socioeconomic factors points to the need for integrated interventions. These interventions should reduce environmental exposures, improve housing quality, strengthen health literacy, and expand access to preventive care. Examples include home-based environmental assessments, community-embedded asthma education programs, mobile health clinics, and workplace or school-based air quality initiatives serving lower-income populations. Including social determinants in asthma action plans and community health worker programs may further help address disparities.

5.4 Clinical Implications

The findings also carry important implications for clinical practice. Healthcare providers caring for adults with asthma may benefit from incorporating routine screening for social determinants of health, including education level and financial barriers, into standard asthma management protocols (Hagan et al., 2021). Clinicians can enhance outcomes by providing tailored patient education, reinforcing inhaler technique, and addressing adherence challenges that may arise from cost or limited health literacy (CDC, 2023). Gender differences observed in this study suggest that clinicians should remain attentive to potential hormonal, behavioral, or psychosocial influences on asthma control among female patients

(Akinbami et al., 2019). Integrating multidisciplinary care models, including respiratory therapists and community health workers, may further strengthen continuity of care and reduce the likelihood of exacerbation.

5.5 Strengths of the Study

This study has several notable strengths, including the use of nationally representative NHANES data, application of complex sampling weights, and inclusion of multiple demographic and clinical predictors. Using recent survey cycles improves the relevance of the findings to contemporary healthcare environments. The multivariable analytical approach also reduces confounding and enhances interpretation of independent associations. Another strength is the study's focus on adult asthma exacerbations rather than prevalence alone. Many national analyses emphasize diagnosis rates or access to care, whereas this investigation prioritized acute attacks as a direct indicator of disease control and healthcare burden. Integrating demographic, behavioral, and clinical predictors within a single multivariable framework allowed for a more nuanced interpretation of independent associations.

5.6 Limitations

This study has several limitations. NHANES relies on self-reported data, which may introduce recall, reporting, or social desirability bias, particularly for behaviors such as smoking or medication adherence. The analytic sample of adults with current asthma was modest, limiting statistical power and precision, and limited variability in certain predictors, particularly health insurance status, constrained the ability to fully assess their associations. Additionally, the cross-sectional nature of the dataset prevents determination of temporal directionality between predictors and outcomes, meaning relationships should be interpreted as correlational rather than causal. Environmental and contextual factors such as air pollution, occupational irritants, and housing quality were not directly measured within the analytic model, restricting the ability to evaluate broader exposure pathways that may interact with socioeconomic variables. These constraints highlight the need for larger longitudinal datasets and more comprehensive environmental measures to provide a fuller understanding of asthma exacerbation risk.

5.7 Future Research Directions

Future research should build upon these findings by employing longitudinal study designs that allow for temporal assessment of insurance continuity and asthma outcomes over time. Incorporating environmental exposure data, medication adherence measures, and geographic variables may provide a more nuanced understanding of exacerbation risk (Liu et al., 2021). Further investigation into interaction effects between socioeconomic factors and biological determinants could also clarify pathways contributing to disparities.

Future investigations may also benefit from incorporating qualitative or mixed methods approaches that explore patient experiences, perceived barriers to care, and cultural influences on medication adherence. Integrating geographic information systems data could illuminate neighborhood-level environmental exposures, transportation accessibility, and healthcare facility distribution. Examining policy-level variables such as Medicaid expansion status or state-specific insurance regulations may further clarify structural influences on asthma control. Expanding research beyond individual predictors toward multilevel models that account for community and policy contexts could yield more actionable insights for reducing asthma disparities nationwide.

Conclusion

In summary, this study contributes to a growing body of evidence demonstrating that asthma exacerbation risk among adults is shaped not only by clinical characteristics but also by broader demographic and socioeconomic factors. The findings indicate that female sex, lower educational attainment, and reduced income are significant predictors of experiencing an asthma attack within a 12-month period. These results reinforce the importance of recognizing asthma as a condition influenced by structural and contextual determinants rather than solely biological mechanisms. Although health insurance status did not reach statistical significance in the adjusted model, the observed direction of association and limited variability within the sample suggest that insurance continuity remains an important area for continued investigation.

From a public health perspective, these findings highlight the need for multifaceted intervention strategies that extend beyond pharmacologic treatment and address upstream determinants of health. Policies that promote equitable access to preventive care, improve housing and environmental conditions, and strengthen health literacy may contribute to improved asthma control and reduced emergency healthcare utilization. Clinically, integrating routine screening for socioeconomic barriers and providing tailored patient education may help mitigate disparities in disease management.

This study also contributes to social science literature by providing updated, nationally representative evidence focused specifically on adult asthma exacerbations. This outcome is examined less frequently than asthma prevalence or access to care alone. By isolating insurance status alongside demographic and behavioral predictors, the analysis offers a clearer understanding of the relative influence of structural healthcare factors. Future research incorporating longitudinal data, environmental exposure measures, and insurance continuity metrics may further clarify causal pathways and inform targeted interventions. Ultimately, advancing asthma management requires coordinated efforts across healthcare systems, community organizations, and policy frameworks to ensure that individuals are not disadvantaged by socioeconomic conditions that impede their ability to maintain respiratory health.

References

Akinbami, L. J., Simon, A. E., & Rossen, L. M. (2019). Changing trends in asthma prevalence among children and adults in the United States. *National Health Statistics Reports*, 123, 1-10. https://pmc.ncbi.nlm.nih.gov/articles/PMC4755484/

Centers for Disease Control and Prevention. (2023). *Asthma in adults*. https://www.cdc.gov/asthma/adults.html

Global Initiative for Asthma. (2022). *Global strategy for asthma management and prevention*. https://ginasthma.org

Grossman, N. L., et al. (2019). Exacerbation-prone asthma in the context of race and ethnic disparities. *Journal of Allergy and Clinical Immunology, 144*(1). https://doi.org/10.1016/j.jaci.2019.03.018

Hagan, J. L., Morley, R. L., & Adams, R. J. (2021). Social determinants of health and adult asthma outcomes: A systematic review. *Journal of Asthma, 58*(10), 1343-1353. https://doi.org/10.1080/02770903.2020.1851302

Henneberger, P. K., Liang, X., Jacobs, D. R., Jr., et al. (2015). Occupational exposures associated with severe exacerbation of asthma among working adults. *American Journal of Industrial Medicine, 58*(6), 695-704. https://doi.org/10.1002/ajim.22435

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387

Liu, A. H., Zeiger, R., & Sorkness, C. A. (2021). Asthma control and disparities in health care access among adults with asthma. *Journal of Allergy and Clinical Immunology, 147*(4), 1340-1350. https://doi.org/10.1016/j.jaci.2020.12.632

Postma, D. S. (2007). Gender differences in asthma development and progression. *Gender Medicine, 4*(Suppl. B), S133-S146. https://doi.org/10.1016/S1550-8579(07)80054-6

Schoen, C., Radley, D., & Riley, P. (2017). Health insurance coverage and health care access among adults with chronic conditions. *Health Affairs, 36*(9), 1541-1549. https://doi.org/10.1377/hlthaff.2017.0360

Smith, A. B., Johnson, T. R., & Lee, M. C. (2021). Health literacy and asthma self-management among adults with chronic respiratory disease. *Journal of Asthma, 58*(7), 921-929. https://doi.org/10.1080/02770903.2020.1731821

Syamlal, G., King, B. A., & Mazurek, J. M. (2022). Prevalence of asthma, asthma attacks, and emergency department visits among working adults in the United States. *Journal of Asthma, 59*(8), 1501-1510. https://doi.org/10.1080/02770903.2021.1901273

Tang, K., Li, S., & Shi, L. (2020). Uninsurance and asthma-related health outcomes in adults. *Journal of Public Health Management and Practice, 26*(5), 453-460. https://doi.org/10.1097/PHH.0000000000001125

Ward, B., Martin, C., & Cooper, J. (2021). Impact of insurance status on asthma management in U.S. adults. *American Journal of Preventive Medicine, 60*(3), 380-388. https://doi.org/10.1016/j.amepre.2020.09.015

World Health Organization. (2023). *Social determinants of health*. https://www.who.int/health-topics/social-determinants-of-health

8. Appendix

8.1 Predictor Variables and NHANES Codes

Table A1.
Predictor Variables and NHANES Codes

NHANES code	Variable or measure
HIQ011	Current health insurance coverage at the time of the interview
SMQ020	Smoking history: ever smoked at least 100 cigarettes
SMQ040	Current smoking status
AGQ030	Hay fever in the past year
BMXBMI	Body mass index (kg/m^2)
DMDEDUC2	Education level
INDFMPIR	Poverty-to-income ratio
RIDAGEYR	Age in years at screening
RIAGENDR	Sex
RIDRETH3	Race and ethnicity
BPQ020	Hypertension
DIQ010	Diabetes status

Evaluating Destination Brand Equity: A Brand Audit of Tirol Through the Lens of Destination Branding Theory

Phani Tej Adidam
University of Nebraska at Omaha

Abstract

Destination brand equity is a key driver of long-term competitiveness in increasingly crowded tourism markets, yet the application of systematic brand audit frameworks to assess the alignment between intended destination identity and consumer perceptions remains limited. This study addresses this gap through a comprehensive destination brand audit of Tirol, Austria, using Konecnik and Gartner's (2007) destination brand equity framework and Keller's (1993) Customer-Based Brand Equity model. A mixed-methods approach combining brand inventory analysis, structured tourist interviews, prospective tourist surveys, and stakeholder consultations was employed to examine four research questions: Tirol's dominant brand associations, the alignment between intended identity and consumer perceptions, brand equity gaps between desired positioning and marketplace reality, and strategies to strengthen competitive positioning in the European Alpine tourism market.

The findings indicate partial alignment between Tirol's "Heart of the Alps" positioning and consumer perceptions. Visiting tourists strongly associate Tirol with nature, alpine sports, and cultural heritage. However, three major brand equity gaps emerged: limited awareness among prospective tourists compared with constituent destinations such as Innsbruck; weak ownership of the "Heart of the Alps" positioning relative to competitors including Graubünden and Salzburg; and fragmented brand architecture that limits the transfer of equity from sub-destinations to the Tirol master brand. Based on these findings, four recommendations are proposed: clarifying Points of Parity and Points of Difference, establishing a unified destination co-branding architecture, implementing integrated multi-channel promotion of the core positioning, and leveraging secondary associations through strategic partnerships and influencer collaborations. This study demonstrates how systematic destination brand audits can diagnose identity–image gaps and generate evidence-based strategies for strengthening destination brand equity in competitive tourism markets.

Keywords: destination branding, brand equity, brand audit, Tirol, Alpine tourism, destination marketing, Customer-Based Brand Equity, identity–image gap

Introduction

Destination branding has emerged as a critical area of research and practice within tourism management, as destinations increasingly compete not only for visitor arrivals but also for attention, meaning, and emotional connection in a crowded global marketplace. Unlike traditional products, destinations are complex, multi-stakeholder entities whose identities are shaped by a diverse range of experiences, narratives, and perceptions that resist simple managerial control (Pike & Page, 2014; Morgan, Pritchard, & Pride, 2011). This complexity makes the development and stewardship of strong destination brands both essential and inherently difficult, as destination marketing organizations (DMOs) must simultaneously satisfy the expectations of visitors, residents, industry partners, and policymakers while projecting a coherent and compelling identity to global audiences.

In response to these challenges, destination marketing scholarship has drawn increasingly upon consumer-based brand equity frameworks to understand how places create and sustain competitive advantage (Konecnik & Gartner, 2007; Boo, Busser, & Baloglu, 2009). Destination brand equity reflects the value generated through the awareness, associations, perceived quality, and emotional connections that consumers develop toward a place (Ruzzier & de Chernatony, 2013; Keller & Swaminathan, 2019). Strong destination brand equity can meaningfully influence travel decision-making, foster repeat visitation and loyalty, support premium pricing, and contribute to sustainable regional development (Blain, Levy, & Ritchie, 2005). Yet destination brands do not exist in a stable environment. Shifting consumer expectations, rapid technological change, intensifying interregional competition, and the disruptions wrought by global events — most recently the COVID-19 pandemic — have compelled destinations to continuously evaluate and adapt their branding strategies. For many established destinations, the central challenge is no longer building a brand from scratch but ensuring that an existing brand remains distinctive, coherent, and resonant across evolving audiences and channels.

It is within this context that brand audits have gained recognition as a vital instrument of destination brand management. A brand audit provides a systematic, evidence-based framework for evaluating both the organizational dimensions of branding — how a destination presents and communicates its identity — and the consumer dimensions — how that identity is actually perceived and interpreted in the marketplace (Adidam & Shaker, 2021; Keller & Swaminathan, 2019). By integrating analysis of brand inventory and brand exploratory components, audits can surface critical gaps between intended positioning and lived consumer perception, identifying strategic vulnerabilities before they erode competitive standing. Despite their established utility in commercial brand management, however, the application of formal brand audit frameworks within destination branding research remains comparatively underexplored, particularly in the context of mature Alpine tourism destinations navigating complex competitive landscapes (Hankinson, 2004; Pike, 2015).

The Austrian region of Tirol[1] provides a compelling and timely context for such an inquiry. Internationally recognized for its alpine landscapes, outdoor recreation, cultural heritage, and world-class tourism infrastructure, Tirol has established itself as one of Europe's preeminent mountain destinations, attracting nearly 47 million overnight stays annually across winter and summer seasons (MCI, 2025). Yet Tirol operates within an intensely competitive Alpine tourism environment that includes formidable

[1] Both “Tirol” and “Tyrol” are used in reference to the Austrian Alpine region. This paper employs “Tirol” when discussing the destination’s official brand identity, as it reflects the spelling used by Tirol Werbung GmbH and the state’s formal communications. “Tyrol” is used selectively in historical or general geographical references.

destinations such as Salzburg, South Tyrol, Graubünden, and Chamonix — all of which are investing heavily in brand differentiation, digital engagement, and experience innovation. As the competitive pressure within this landscape increases, maintaining a distinctive, coherent, and emotionally resonant destination identity has become a strategic imperative rather than a peripheral concern (Buhalis, 2000).

Despite Tirol's prominence, limited academic research has examined the degree to which its intended brand identity aligns with the perceptions held by visitors and prospective travelers. In particular, there is a need to better understand the extent to which Tirol's core brand positioning — encapsulated in its mantra "Heart of the Alps" — resonates with diverse target audiences (see Figure 1), the nature and strength of the associations consumers link to the destination, and the strategic implications of any gaps between the brand Tirol projects and the brand consumers actually experience. Without this diagnostic understanding, branding investments risk being misallocated and strategic decisions risk being made in the absence of consumer reality.

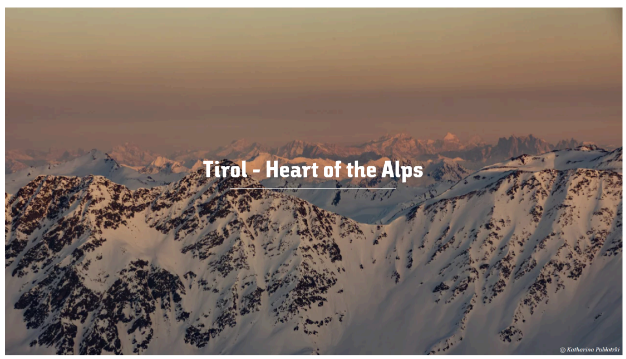


Figure 1. Brand Mantra of Tirol (Source: Tirol Werbung GmbH, 2025)

To address this gap, the present study conducts a comprehensive destination brand audit of Tirol, drawing upon brand equity theory and destination branding literature to integrate brand inventory analysis, tourist surveys, and stakeholder interviews. The study is guided by the following research questions:

- **RQ1:** What are the dominant brand associations, images, and perceptions — both cognitive and affective — held by current and prospective tourists regarding Tirol as an Alpine destination?
- **RQ2:** To what extent does Tirol's intended brand identity, as expressed through its "Heart of the Alps" positioning and official brand communications, align with the associations and perceptions that consumers actually hold?

- **RQ3:** What gaps exist between Tirol's desired brand positioning and its actual levels of brand awareness, brand mantra ownership, and brand architecture coherence among visiting and prospective tourist segments?
- **RQ4:** What evidence-based destination branding strategies, grounded in the audit findings, can most effectively strengthen Tirol's brand feelings, competitive differentiation, and overall brand equity within the European Alpine tourism market?

This study contributes to the destination branding literature in three interconnected ways. First, it demonstrates the systematic application of a two-stage brand audit framework — combining brand inventory and brand exploratory analysis — within the specific context of destination marketing, responding to scholarly calls for more holistic, mixed-methods approaches to evaluating destination brand equity (Hankinson, 2004; Hanna & Rowley, 2015). Second, it generates empirical insights into the relationship between intended destination identity and consumer-based brand perceptions, illuminating the gap dynamics that shape competitive vulnerability in mature Alpine destinations. Third, it translates audit findings into evidence-based strategic recommendations grounded in destination branding theory, offering a replicable model that DMOs and policymakers can adapt to their own contexts.

Through the case of Tirol, this study illustrates a broader principle: that destination brand audits are not merely evaluative exercises but generative strategic processes — tools through which destinations can move from assumption to evidence, from fragmentation to coherence, and from being merely visited to being genuinely valued.

Literature Review and Theoretical Foundation

Destination Branding: Conceptual Foundations

The study of destination branding sits at the intersection of place marketing, consumer psychology, and strategic management, and has grown substantially over the past three decades as tourism markets have become more competitive and travelers more discerning. Early destination marketing efforts focused largely on promotional activities — advertising campaigns, brochures, and trade fair participation — designed to increase visitor awareness and arrivals. However, as Pike and Page (2014) observe in their comprehensive narrative review of the field, scholars and practitioners gradually recognized that sustainable destination competitiveness required something deeper than promotion: it required the construction and management of a coherent, differentiated, and emotionally resonant brand identity.

This conceptual shift drew heavily on commercial branding literature. Aaker's (1991) foundational work on brand equity and Keller's (1993) customer-based brand equity framework provided the theoretical scaffolding upon which destination branding scholars began to build destination-specific models. Morgan, Pritchard, and Pride (2011) argued that in an era of information overload and global destination proliferation, places that fail to brand themselves strategically risk becoming commodities — indistinguishable from competitors and unable to command loyalty or premium positioning. This insight gave rise to a growing body of destination branding scholarship that sought to understand not merely how destinations could be promoted, but how they could be *branded* — that is, how they could create distinctive, emotionally meaningful identities that endure beyond individual marketing campaigns.

Destination branding has since been defined as the process of creating a unique and compelling identity that differentiates a place from competing destinations while fostering favorable and enduring perceptions among target audiences (Cai, 2002). Effective destination brands communicate a clear value proposition, reduce perceived travel risk, facilitate recognition during the travel planning process, and enhance post-visit loyalty (Blain, Levy, & Ritchie, 2005). Research has consistently demonstrated that destinations with strong brands are more likely to attract visitors, generate positive word-of-mouth, secure media attention, and maintain long-term competitiveness in crowded tourism markets (Pike, 2015).

A defining challenge in destination branding, and one that distinguishes it from commercial product branding, is the inherent complexity of the destination as a brand entity. Unlike a manufactured product whose attributes can be controlled and standardized, a destination is a multidimensional entity involving multiple stakeholders — including tourism boards, government agencies, hospitality providers, residents, and visitors themselves — each of whom may hold different understandings of what the destination represents (Blain, Levy, & Ritchie, 2005). The destination experience is co-created through countless interactions and touchpoints that no single organization can fully control or choreograph (Hanna & Rowley, 2015). Moreover, consumer perceptions of a destination are shaped not only by official brand communications but also by personal experiences, peer recommendations, social media content, news coverage, and cultural representations in film and literature (Xiang, Du, Ma, & Fan, 2017). Consequently, achieving alignment between a destination's intended brand identity and its perceived brand image among consumers represents one of the central — and most persistently difficult — challenges in destination marketing.

It is precisely this alignment challenge that motivates the present study. As established in the introduction, Tirol's "Heart of the Alps" positioning represents a deliberately constructed brand identity; the question this audit investigates is the degree to which that identity resonates in the minds of those it seeks to reach.

Destination Brand Equity: A Multi-Dimensional Framework

The concept of brand equity offers a powerful theoretical lens through which to evaluate destination brand strength. In its foundational commercial sense, brand equity refers to the differential effect that brand knowledge has on consumer responses to marketing activities (Keller & Swaminathan, 2019). Applied to tourism, destination brand equity reflects the awareness, associations, perceived quality, and emotional connections that collectively influence destination choice, visitor behavior, and long-term loyalty (Konecnik & Gartner, 2007).

Konecnik and Gartner's (2007) seminal adaptation of brand equity theory to the destination context established a multi-dimensional model comprising four interrelated components: destination awareness, destination image, perceived destination quality, and destination loyalty. This model, subsequently refined and validated across multiple destination contexts by Boo, Busser, and Baloglu (2009), provides a robust framework for measuring and interpreting the sources of destination brand value. Each dimension contributes distinctively to overall equity, and weaknesses in any one dimension can undermine the strength of the others.

Destination awareness represents the extent to which consumers recognize or recall a destination when prompted or when considering travel options. Awareness is widely regarded as the foundational dimension of destination brand equity: a destination cannot be chosen if it is not known (Pike, 2015). High awareness increases the probability that a destination will be included in a traveler's consideration set during the decision-making process, and it also amplifies the effectiveness of other brand equity dimensions — associations, image, and quality — by providing a cognitive anchor to which these attributes can attach. Critically, awareness operates at multiple levels: unaided recall (spontaneously naming a destination) differs meaningfully from aided recognition (confirming familiarity when prompted), and both differ from top-of-mind salience, which reflects the strength and accessibility of the brand in memory (Keller & Swaminathan, 2019). This awareness asymmetry — well documented in multi-level destination contexts — is a central concern the present audit investigates.

Destination image refers to the set of beliefs, impressions, and associations that individuals hold regarding a place. These associations encompass both cognitive components — functional attributes such as scenery, climate, infrastructure, and accessibility — and affective components, including the emotional responses the destination evokes, such as excitement, tranquility, authenticity, or adventure (Baloglu & McCleary, 1999). Research has consistently shown that positive and distinctive destination images significantly predict destination preference and intention to visit, and that affective image components often exert stronger influence on behavioral intentions than cognitive components alone (Pike & Page, 2014). The composition of destination image is particularly consequential for competitive differentiation: where competing destinations share similar cognitive attributes — as is the case across the Alpine tourism landscape — affective image differentiation becomes the primary driver of preference.

Perceived destination quality reflects visitors' evaluations of the destination's ability to deliver satisfying, high-value experiences. Hospitality, service standards, safety, infrastructure quality, environmental integrity, and value for money collectively shape these perceptions (Konecnik & Gartner, 2007). Quality perceptions serve as credibility signals that reinforce or undermine the brand's other associations: a destination that projects an image of authenticity and natural beauty but delivers inconsistent service quality risks eroding the very associations it seeks to cultivate. In the Alpine context, where competing destinations generally operate at high quality standards, perceived quality differentials may be subtle but nonetheless consequential for loyalty and premium positioning.

Destination loyalty encompasses both repeat visitation intentions and positive word-of-mouth advocacy (Pike & Page, 2014). Loyal visitors function as informal brand ambassadors, contributing to organic destination promotion through recommendations, social media sharing, and peer influence — channels that research increasingly identifies as more persuasive than formal advertising, particularly among younger travelers (Hudson et al., 2015). Loyalty is not merely a behavioral outcome but a relational one: it reflects the depth of psychological attachment that consumers develop toward a destination over time, and it is this depth of connection — more than any single visit — that sustains destination brand equity across competitive cycles (Boo, Busser, & Baloglu, 2009).

Keller's Customer-Based Brand Equity Framework and Its Application to Destinations

Among the theoretical frameworks informing destination brand equity research, Keller's Customer-Based Brand Equity (CBBE) model remains one of the most widely applied and adapted (Keller & Swaminathan, 2019). The CBBE framework proposes that strong brands are built by establishing deep, favorable, and unique associations in consumers' minds, and that these associations develop sequentially through four hierarchical stages: brand identity, brand meaning, brand responses, and brand relationships.

At the model's foundation lies **brand identity**, which is concerned with brand salience — the breadth and depth of brand awareness, or how readily and consistently the brand comes to mind across different consumption contexts. For destinations, salience reflects not only whether travelers have heard of a place but whether it is recalled in relevant decision moments, such as planning a skiing holiday or an Alpine cultural trip. The second stage, **brand meaning**, encompasses both performance-related associations (the functional qualities of the destination experience) and imagery-related associations (the symbolic, social, and emotional meanings the destination conveys). In Tirol's case, performance associations include its world-class ski infrastructure, accessibility from major European cities, and hospitality quality, while imagery associations encompass Tyrolean cultural heritage, alpine authenticity, and the "Heart of the Alps" identity.

The third stage, **brand responses**, captures consumer judgments and feelings toward the brand. Judgments reflect rational evaluations of quality, credibility, and superiority relative to alternatives, while feelings represent emotional reactions — excitement, warmth, adventure, self-respect — that the brand evokes. Research in tourism contexts confirms that emotional brand responses exert disproportionate influence on behavioral intentions, particularly revisit intentions and recommendation behaviors (Pike & Page, 2014). The study's finding that 42% of visiting tourists associated Tirol primarily with excitement underscores the affective richness of the destination experience and its potential as a brand positioning asset. The final stage, **brand resonance**, reflects the quality of the consumer-brand relationship — the degree of active loyalty, attachment, community, and engagement that consumers sustain with the brand over time.

The CBBE model is particularly valuable for this study because it provides a structured, theoretically grounded vocabulary for interpreting audit findings across multiple brand equity dimensions. Rather than treating brand strength as a single undifferentiated construct, the framework allows for a nuanced diagnosis of where equity is strong, where it is weak, and where strategic interventions are most likely to produce meaningful gains.

The Identity–Image Gap: A Central Problem in Destination Branding

A recurrent theme across destination branding scholarship is the gap between a destination's intended brand identity — the associations and meanings its managers seek to project — and its actual brand image — the associations and meanings that consumers actually hold (Hankinson, 2004; Mariani, Di Felice, & Mura, 2016; Ruzzier & de Chernatony, 2013). This identity–image gap is theoretically significant because it reveals the limits of managerial control over destination brand meaning and highlights the mediating role of consumer experience, media interpretation, and competitive context.

Hankinson (2004) conceptualizes destination brand identity as a relational network of associations radiating outward from a core brand positioning statement, mediated by communications, infrastructure, and service experiences. Where this relational network is coherent and consistently reinforced, consumer image tends to converge with intended identity. Where it is fragmented — as when a destination's marketing communicates one message while its sub-regions communicate another, or when individual resorts attract stronger recognition than the parent brand — the identity–image gap widens, and brand equity is correspondingly weakened.

This fragmentation dynamic is directly observable in Tirol's case and motivates several of the strategic recommendations developed later in this paper. When visitors recognize Kitzbühel, Innsbruck, and Ischgl as individual destinations but fail to associate them with Tirol as an overarching brand, the equity generated through positive resort experiences does not transfer to the parent brand — a pattern Blain, Levy, and Ritchie (2005) describe as a failure of brand architecture rather than a failure of destination quality. Addressing this gap requires not merely more promotion but a more coherent brand architecture that connects sub-destination experiences to the master brand in ways that are visible, consistent, and emotionally meaningful.

Brand Audits as Strategic Assessment Tools

Brand audits represent one of the most systematic and evidence-based approaches to diagnosing the identity–image gap and evaluating overall brand equity. Adidam and Shaker (2021) define a brand audit as a comprehensive, systematic examination of a brand's current strengths, weaknesses, and strategic opportunities, conducted from both organizational and consumer perspectives. Unlike a marketing audit, which evaluates the effectiveness of specific promotional activities, a brand audit assesses the holistic health of the brand as an asset — examining whether it is positioned correctly, perceived accurately, and managed coherently across all touchpoints.

Brand audits are structured around two complementary components. The **brand inventory** examines the organization's own perspective: the brand elements (name, logo, tagline, imagery), communications strategies, partnerships, pricing signals, and positioning statements that collectively constitute the destination's intended brand identity. This component is descriptive and managerially focused, providing a comprehensive account of how the destination seeks to present itself. The **brand exploratory** examines the consumer perspective through qualitative and quantitative research methods — surveys, interviews, focus groups, and perceptual mapping — to uncover the awareness levels, associations, attitudes, emotional responses, and loyalty intentions that consumers actually hold (Keller & Swaminathan, 2019). Together, these components generate a diagnostic portrait of the brand's current position in the marketplace and illuminate the specific points at which intended identity and perceived image diverge.

Within destination branding specifically, brand audits have been identified as particularly valuable tools because they surface discrepancies that routine performance metrics — visitor numbers, occupancy rates, social media engagement — cannot capture (Pike, 2015). A destination may achieve strong arrival figures while simultaneously suffering from low brand salience, weak differentiation, or a narrowly defined image that limits its appeal across seasons or audience segments. Audits make these structural vulnerabilities visible, providing evidence-based foundations for strategic repositioning rather than relying on managerial assumption alone.

Despite this recognized utility, the systematic application of brand audit frameworks within destination branding research remains comparatively limited, particularly for mature Alpine destinations navigating competitive repositioning (Hankinson, 2004; Hanna & Rowley, 2015). Most existing empirical studies focus on a single dimension of destination brand equity — typically image or awareness — rather than adopting the integrative, multi-component approach that a full brand audit provides. This study responds to that gap by applying a comprehensive brand audit framework to Tirol, combining brand inventory analysis, tourist surveys, and stakeholder interviews to generate a holistic evaluation of destination brand equity.

Conceptual Framework

Drawing upon the theoretical foundations established above — destination branding theory, Konecnik and Gartner's (2007) multi-dimensional brand equity model, and Keller's Customer-Based Brand Equity framework — this study proposes a conceptual framework in which destination brand equity is shaped by the dynamic interaction between a destination's intended brand identity and the perceptions, associations, and emotional responses that consumers actually hold. This interaction is evaluated across four interconnected dimensions, each of which corresponds directly to one of the study's research questions and provides the analytical structure through which the brand audit findings and strategic recommendations are organized.

Dimension 1: Brand Associations and Image captures the cognitive and affective meanings that current and prospective tourists attach to Tirol. This dimension encompasses both the functional attributes consumers associate with the destination — natural landscapes, alpine sports, cultural heritage — and the symbolic and emotional meanings these associations convey. It reflects Keller's brand meaning stage of equity development and corresponds to the study's first research question concerning the dominant associations and perceptions tourists hold regarding Tirol.

Dimension 2: Brand Identity Alignment examines the degree of correspondence between Tirol's intended brand identity — as expressed through its "Heart of the Alps" mantra, visual brand elements, and official communications — and the perceptions and associations that consumers spontaneously generate. Drawing on Hankinson's (2004) relational network conceptualization of destination brand identity and the identity–image gap literature, this dimension evaluates whether Tirol's strategic positioning is achieving the consumer resonance it seeks. It corresponds to the study's second research question concerning the alignment between intended identity and consumer perceptions.

Dimension 3: Brand Awareness and Equity Gaps assesses the extent to which consumers recognize and recall Tirol across aided and unaided conditions, examines how awareness levels vary between visiting and prospective tourist segments, and identifies the structural gaps — in awareness, mantra ownership, and brand architecture — that constrain the accumulation of destination brand equity. Consistent with Konecnik and Gartner's (2007) positioning of awareness as the foundational prerequisite of destination brand equity, this dimension corresponds to the study's third research question concerning the gaps between Tirol's desired positioning and its actual marketplace presence.

Dimension 4: Brand Feelings, Competitive Positioning, and Strategic Implications integrates the emotional responses generated by the destination experience — including feelings of excitement, adventure, and fun — with an assessment of how Tirol differentiates itself from competing Alpine destinations in consumer decision-making. By connecting affective brand responses to competitive positioning, this dimension identifies the strategic opportunities most likely to strengthen Tirol's brand equity and provides the empirical foundation for the study's evidence-based recommendations. It corresponds to the study's fourth research question concerning the destination branding strategies best suited to enhance Tirol's competitive standing.

Together, these four dimensions move the study beyond descriptive brand assessment toward a diagnostic and strategic evaluation of Tirol's destination brand equity — one that is simultaneously grounded in established theory, anchored in empirical evidence, and oriented toward actionable managerial implications.

Drawing upon destination branding theory, Keller's Customer-Based Brand Equity framework, and the destination brand equity literature, Figure 2 presents the conceptual framework guiding this study. The framework illustrates how Tirol's intended brand identity is evaluated through a brand audit process that examines consumer awareness, brand associations, emotional responses, and competitive positioning, ultimately revealing brand equity gaps and informing strategic recommendations.

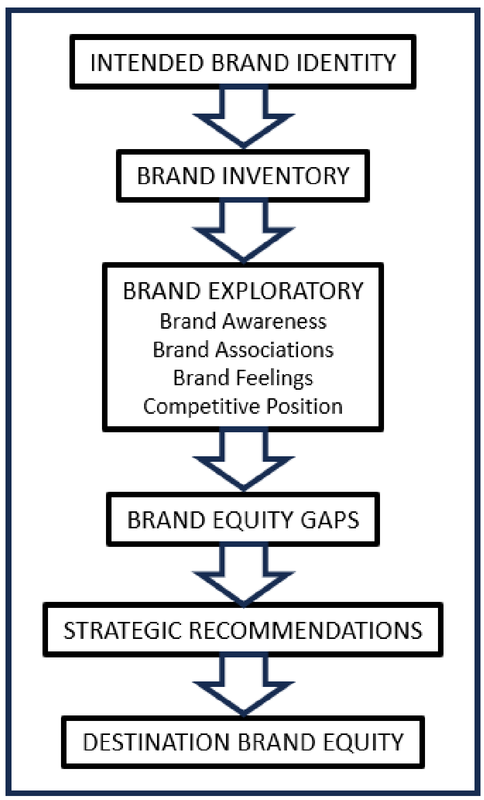


Figure 2. Conceptual Framework for Evaluating Destination Brand Equity

Research Methodology

Research Design

To address the research questions outlined in the introduction, this study employed a mixed-methods brand audit approach integrating qualitative and quantitative research techniques. Consistent with established brand audit frameworks (Adidam & Shaker, 2021; Keller & Swaminathan, 2019), the study was structured around two complementary components: a brand inventory and a brand exploratory. The brand inventory examined Tirol's intended brand identity through systematic analysis of official marketing materials, digital communications, brand elements, and stakeholder interviews. The brand exploratory investigated consumer perceptions, awareness levels, emotional responses, and destination associations through structured personal interviews and questionnaire surveys.

The mixed-methods design was selected for two reasons. First, it enables triangulation across multiple data sources — organizational communications, consumer surveys, tourist interviews, and industry stakeholder perspectives — thereby strengthening the credibility and robustness of the findings. Second, it reflects the inherently dual nature of brand audit inquiry, which requires both an inside-out perspective (how the destination presents itself) and an outside-in perspective (how consumers actually perceive it). The alignment — or misalignment — between these two perspectives constitutes the central diagnostic output of the audit (Adidam & Shaker, 2021).

Brand Inventory

The brand inventory was constructed through a systematic review of Tirol's branding and communication activities across digital and print channels. Specifically, the analysis encompassed Tirol's official tourism websites, social media platforms, destination brochures, promotional collateral, advertising materials, licensing and merchandising programs, and published brand guidelines. These materials were examined to identify how Tirol presents its brand mantra, value proposition, visual identity, and target market positioning across consumer-facing touchpoints.

To complement the documentary analysis, semi-structured interviews were conducted with key representatives of Tirol Werbung GmbH, Tirol's official destination marketing organization. Interview guides were structured around brand inventory categories drawn from Adidam and Shaker's (2021) brand audit framework, covering branding objectives, communication strategies, partnership activities, and positioning priorities over the preceding five years. These interviews provided an organizational perspective on the intended brand identity that the documentary review alone could not fully capture, particularly regarding strategic intentions that may not be fully visible in public-facing communications.

Brand Exploratory Instrument Development

The brand exploratory instruments were developed using established destination branding and brand equity constructs drawn from Konecnik and Gartner's (2007) multi-dimensional destination brand equity model and Keller's Customer-Based Brand Equity (CBBE) framework (Keller & Swaminathan, 2019). The questionnaire consisted of 18 items measuring destination awareness, brand associations, destination image, emotional responses, and competitive positioning, while the interview guides were structured around the same theoretical dimensions. Prior to data collection, the instruments were reviewed by two tourism academics and pilot-tested with ten international travelers to assess clarity, face validity, and alignment with the study's research objectives. This process helped ensure that the data collected were theoretically grounded, analytically relevant to the research questions, and comparable with prior destination branding research.

Tourist Interviews

Structured personal interviews were conducted with 114 tourists currently visiting Tirol. Respondents were approached at three high-traffic tourism locations: Innsbruck Airport, Innsbruck Railway Station, and the Bergisel tourism area. These sites were selected purposively because they represent distinct visitor convergence points — arriving, departing, and sightseeing tourists respectively — thereby maximizing demographic and motivational diversity within the sample while remaining practically accessible for systematic data collection. The interview instrument explored destination awareness, spontaneous brand associations, emotional responses to the destination experience, competitive destination considerations, and perceptions of Tirol's overall positioning relative to other Alpine destinations.

Survey of Prospective Tourists

A structured questionnaire was distributed to 300 prospective tourists across ten countries representing Tirol's primary and emerging international source markets, as identified in Tirol Werbung's own market intelligence. Critically, the survey targeted individuals with no prior travel experience in Tirol, ensuring that responses reflected external destination perceptions formed independently of firsthand visits. This design choice was deliberate: by capturing perceptions among non-visitors, the study could assess Tirol's brand reach and the effectiveness of its communications beyond its existing visitor base — a dimension of brand equity that in-destination surveys alone cannot provide. Of the 300 questionnaires distributed, 120 usable responses were returned, yielding a response rate of 40%.

Stakeholder Interviews

To supplement consumer perspectives with industry viewpoints, interviews were conducted with sixteen tourism operators and travel agency representatives across multiple international markets. Five operators were based in Innsbruck, providing a local industry perspective, while eleven were located across major international tourism hubs — Vienna, Paris, Munich, Venice, Rome, and New Delhi — offering insight into how Tirol is positioned, perceived, and sold within key source markets. These stakeholder interviews provided a third analytical perspective on Tirol's brand positioning, complementing the organizational view captured through the brand inventory and the consumer view captured through tourist interviews and surveys.

Data Analysis

Quantitative data from the prospective tourist survey were analyzed using descriptive statistical techniques, including frequency distributions and percentage analyses. These methods were employed to identify patterns in destination awareness, brand associations, emotional responses, and competitor evaluations across the foreign tourist sample.

Qualitative data generated through tourist interviews and stakeholder discussions were analyzed using thematic analysis. Interview responses were systematically coded and categorized according to the destination branding constructs identified in the literature review — specifically destination awareness, brand image, brand associations, emotional responses, and competitive positioning — thereby maintaining consistency between the theoretical framework and the analytical categories applied to the data. In the case of tourist interviews, association mapping was employed to identify clusters of cognitive and affective meanings linked to the Tirol brand, the results of which are visualized in the mental map presented in the findings section.

The integration of quantitative survey findings, qualitative interview themes, and documentary brand inventory analysis facilitated cross-source triangulation, enabling the research team to identify convergent patterns as well as points of tension between organizational intentions and consumer perceptions.

Validity, Reliability, and Limitations

Several measures were adopted to strengthen the validity and credibility of the study's findings. First, data triangulation was achieved by collecting evidence from three distinct respondent groups — current visitors, prospective visitors, and industry stakeholders — as well as from organizational documentation, reducing reliance on any single source and enabling cross-validation of key findings. Second, the alignment of survey and interview instruments with established theoretical constructs from the destination brand equity literature provided a degree of construct validity, ensuring that the dimensions measured were conceptually grounded and consistent with prior research. Third, the combination of qualitative and quantitative methods allowed complementary dimensions of destination brand perception to be captured, with qualitative findings providing contextual depth and quantitative findings providing distributional breadth.

The study's limitations should also be acknowledged. The use of convenience sampling at fixed tourist locations, while practical and common in destination research, limits the external validity of findings and may introduce sampling bias toward visitors who frequent particular site types. The cross-sectional research design captures perceptions at a single point in time and cannot account for seasonal variation in brand associations or the dynamic effects of marketing campaigns and external events on consumer perceptions. Additionally, self-reported brand perceptions and awareness measures are subject to recall bias and social desirability effects, which may lead some respondents to overstate familiarity with destination brands. Finally, while the prospective tourist sample spanned ten countries, the relatively modest sample size of 120 respondents limits the granularity of cross-national comparisons. Notwithstanding these limitations, the mixed-methods design and multi-source triangulation approach provide a credible and analytically robust foundation for evaluating Tirol's destination brand equity and deriving strategic implications for destination management.

Findings

The findings are organized according to the four research questions guiding the study, integrating evidence from tourist interviews, prospective tourist surveys, stakeholder interviews, and brand inventory analysis. Together, they provide a diagnostic portrait of Tirol's current destination brand equity — its strengths, its gaps, and the points at which intended identity and consumer perception diverge most consequentially.

RQ1: What are the dominant brand associations, images, and perceptions — both cognitive and affective — held by current and prospective tourists regarding Tirol as an Alpine destination?

Tourist interviews and mental mapping exercises revealed that current visitors associate Tirol primarily with three overarching thematic clusters: nature, alpine sports and adventure, and tradition and culture. These clusters emerged consistently and independently across respondents, suggesting that they represent robust, widely shared cognitive associations rather than idiosyncratic individual perceptions. Within the nature cluster, respondents referenced mountains, rivers, lakes, animals, and forests. Alpine sports associations encompassed skiing, snowboarding, ski jumping, biking, climbing, and the Winter Olympic Games. The tradition and culture cluster included references to Tyrolean language and customs, historic sightseeing landmarks such as the Goldenes Dachl, and cultural characteristics perceived as distinctive to the region. Additional associations spanning cuisine, lifestyle, sustainability, and family heritage further enriched the mental map, suggesting that Tirol's brand meaning extends across multiple experiential domains (see Figure 3).

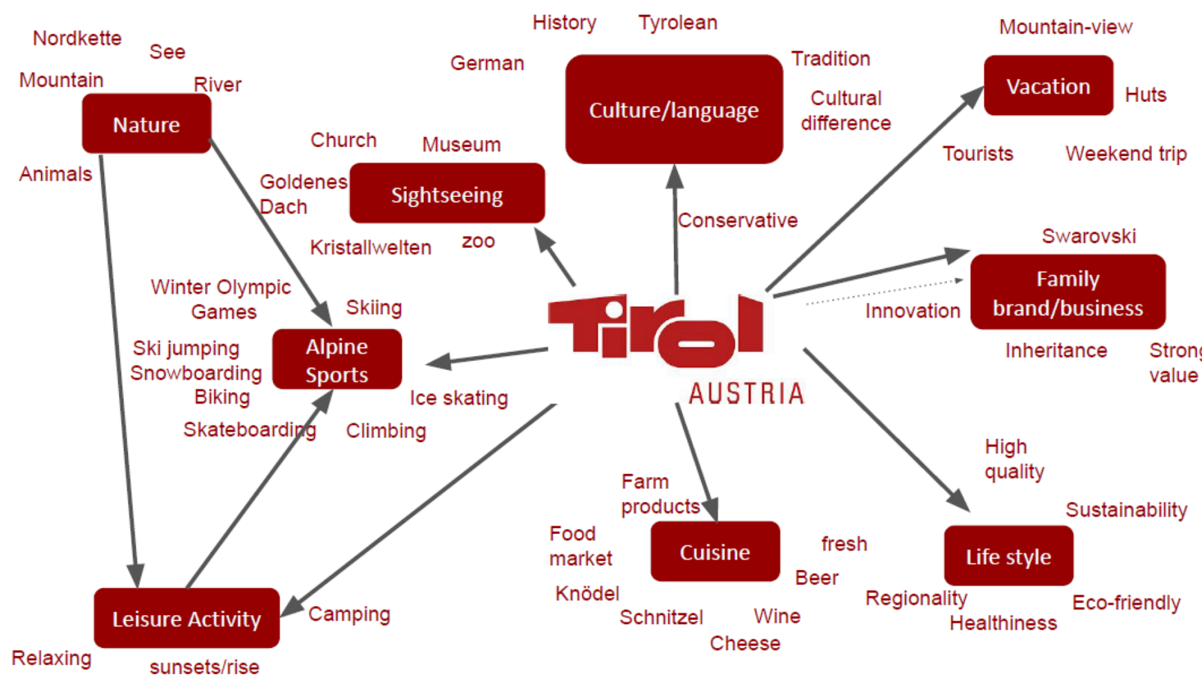


Figure 3. Mental Map of Tirol (based on surveys of 114 Tyrol tourists)

These cognitive associations were accompanied by strongly positive emotional responses. Excitement emerged as the dominant brand feeling among visiting tourists, cited by 42% of respondents, followed by fun (22%) and adventure (20%). Smaller proportions of respondents reported feelings of social approval, self-respect, warmth, and confidence (see Table 1). Collectively, these findings indicate that Tirol generates a rich and emotionally resonant visitor experience. The concentration of emotional responses around excitement, fun, and adventure is particularly significant from a brand equity perspective: consistent with Keller's CBBE framework, these affective associations represent the brand feeling stage of equity development, reflecting emotional responses that the destination reliably evokes and that can, when leveraged strategically, deepen into the brand resonance and loyalty that sustain long-term competitiveness.

Stakeholder interviews with tourism operators across Innsbruck and international markets broadly corroborated these consumer findings, with operators frequently characterizing Tirol as an authentic, high-quality Alpine destination with particular strength in winter sports and outdoor recreation. Several respondents emphasized the destination's reputation for reliability, infrastructure quality, and year-round outdoor experiences. As one Innsbruck-based operator observed, *"Visitors consistently associate Tirol with skiing, mountain landscapes, and a genuine Alpine atmosphere."* Similarly, a Munich-based travel representative noted that *"Tirol is viewed as one of the benchmark Alpine destinations, particularly for winter sports and active holidays."*

Table 1. Visiting Tyrol, gives you a *feeling* of ______ ?

Excitement	41
Fun	22
Adventure	20
Social Approval	11
Self-respect	9
Warmth	5
Confidence	4
Others	2
	114

However, several international operators—particularly those based in Rome, Paris, and New Delhi—indicated that while individual resorts such as Kitzbühel and Innsbruck were well recognized among their clients, the Tirol brand as a regional entity was considerably less familiar. One operator remarked that *"many travelers know Innsbruck immediately, but do not necessarily realize that it is part of Tirol."* Another noted that *"clients often ask specifically for Kitzbühel or St. Anton rather than for Tirol as a destination."* These recurring observations suggest that awareness is concentrated at the resort and city level rather than at the regional brand level, a pattern that anticipates the awareness and brand architecture gaps discussed under RQ3.

RQ2: To what extent does Tirol's intended brand identity, as expressed through its "Heart of the Alps" positioning and official brand communications, align with the associations and perceptions that consumers actually hold?

The brand inventory established that Tirol's intended brand identity is anchored in the mantra "Heart of the Alps," communicating a positioning built around alpine lifestyle, sustainability, outdoor recreation, cultural heritage, and holistic well-being. Tirol Werbung's communications consistently project a multifaceted destination capable of delivering experiences across sports, family, wellness, and cultural tourism segments, united by a coherent regional identity.

The findings reveal a pattern of **partial alignment** between this intended identity and consumer perceptions. On one hand, the dominant associations reported by visiting tourists — nature, sports and adventure, tradition and culture — resonate meaningfully with the core elements of the "Heart of the Alps" positioning. Visitors perceive Tirol as authentically Alpine, experientially rich, and culturally distinctive, which aligns well with the destination's strategic intent. The emotional associations of excitement and adventure further reinforce the active, aspirational dimensions of the brand that Tirol Werbung seeks to project.

Table 2. Which destination comes to mind when you think of “Heart of the Alps”?

	Tyrol Tourists (%)	**Foreign Tourists (%)**
Graubünden (Switzerland)	23	37
Salzburg (Austria)	29	32
Tirol	22	3
Innsbruck	7	13
Chamonix (France)	12	13
Others	7	2

On the other hand, the alignment breaks down at the level of explicit brand positioning. When visiting tourists were asked to associate the phrase "Heart of the Alps" with a specific destination, only 22% identified Tirol — placing it third behind Salzburg (29%) and Graubünden (23%). Among prospective tourists with no prior Tirol experience, the misalignment was markedly more pronounced: only 3% associated "Heart of the Alps" with Tirol, compared to 37% for Graubünden and 32% for Salzburg (see Table 2). Innsbruck — Tirol's own capital city — was associated with the mantra four times more frequently than Tirol itself among the prospective tourist sample.

These findings indicate that while Tirol's experiential and associative brand content is broadly consistent with its intended identity, the overarching brand mantra has not achieved sufficient marketplace ownership to function as an effective differentiator. Consumers are experiencing the destination in ways that largely reflect its intended values, but they are not connecting those experiences to the "Heart of the Alps" positioning in meaningful numbers. This represents a failure not of destination quality but of brand communication: the connective tissue between experience and identity is insufficiently developed. As Hankinson (2004) and Ruzzier and de Chernatony (2013) suggest, this type of identity–image gap is among the most strategically consequential challenges a destination brand can face, because it means that positive consumer experiences are generating equity for the category — Alpine tourism broadly — rather than for the specific destination brand.

RQ3: What gaps exist between Tirol's desired brand positioning and its actual levels of brand awareness, brand mantra ownership, and brand architecture coherence among visiting and prospective tourist segments?

The findings reveal three significant and interrelated brand equity gaps, each with distinct strategic implications.

Gap 1: Limited Brand Awareness Among Prospective Tourists

Brand awareness data exposed a substantial gap between Tirol's recognition among current visitors and its reach among prospective tourists. While 66.7% of tourists interviewed within Tirol reported familiarity with the Tirol brand, this figure dropped sharply to 13.3% among prospective tourists who had not previously visited the region (see Table 3a). The contrast with Innsbruck is particularly striking: 100% of in-destination tourists and 88.3% of prospective tourists were familiar with Innsbruck (see Table 3b). This awareness asymmetry is theoretically significant because, as Konecnik and Gartner (2007) establish, awareness functions as the foundational prerequisite of destination brand equity — a destination that is not known cannot be considered, chosen, or recommended. Tirol's comparatively low awareness among non-visitors suggests that a substantial proportion of its potential audience is effectively unreachable through brand associations alone, regardless of the strength of those associations among existing visitors.

Table 3a. Have you heard of Tirol?

	Tyrol Tourists	Foreign Tourists
Yes	66.70%	13.30%
No	33.30%	86.70%
	114	120

Table 3b. Have you heard of Innsbruck?

	Tyrol Tourists	Foreign Tourists
Yes	100.00%	88.30%
No	0.00%	11.70%
	114	120

Gap 2: Weak Marketplace Ownership of the Brand Mantra

As documented under RQ2, the "Heart of the Alps" mantra demonstrates limited ownership in the competitive Alpine tourism landscape. The finding that competing destinations — most notably Graubünden and Salzburg — are more frequently associated with this concept than Tirol itself, among both current and prospective tourists, indicates that the mantra has not successfully differentiated Tirol's positioning in the consumer mind. From a CBBE perspective, this represents a weakness at the brand meaning stage of equity development: the destination's core positioning statement is not generating the strong, favorable, and unique associations that Keller and Swaminathan (2019) identify as the building blocks of competitive brand advantage. A brand mantra that consumers attribute more readily to competitors than to the brand itself is, by definition, not functioning as a point of difference.

Gap 3: Fragmented Brand Architecture

A third gap concerns the structural relationship between Tirol's individual resort destinations and the overarching regional brand. The findings indicate that tourists frequently develop strong associations with specific destinations — Kitzbühel, Innsbruck, Ischgl, Zillertal, and St. Anton among the most cited — without necessarily connecting these experiences to the broader Tirol identity. When asked which destinations they had considered as alternatives before choosing Tirol, respondents cited South Tyrol (39%) and Salzburg (35%) most frequently, suggesting that competitive consideration occurs primarily at the sub-destination level rather than the regional brand level (see Table 4). This pattern of fragmentation — in which sub-destination equity fails to transfer upward to the parent brand — is consistent with what Blain, Levy, and Ritchie (2005) describe as a brand architecture failure: the equity generated through positive individual resort experiences remains anchored at the resort level rather than accumulating within the Tirol master brand. The net effect is that Tirol's considerable tourism infrastructure and visitor satisfaction levels are not translating into proportionate regional brand equity.

Table 4. What alternate destination did you consider?

South Tirol	39
Salzburg	35
Chamonix	14
Graubünden	11
Vorarlberg	6
Carinthia	5
Others	4
	114

RQ4: What evidence-based destination branding strategies, grounded in the audit findings, can most effectively strengthen Tirol's brand feelings, competitive differentiation, and overall brand equity within the European Alpine tourism market?

The convergent findings across all three equity gaps point toward four interconnected strategic priorities. First, Tirol requires a more deliberate and sustained effort to build brand awareness among prospective tourist segments in key international source markets, with particular attention to audiences for whom Innsbruck is familiar but Tirol is not — a group that represents both a significant gap and a significant opportunity. Second, the "Heart of the Alps" mantra requires more aggressive and consistent multi-channel promotion if it is to achieve the marketplace ownership necessary to function as a genuine differentiator. Third, the brand architecture gap between individual resorts and the Tirol master brand must be addressed through structural co-branding mechanisms that make the regional identity visibly present at individual destination touchpoints. Fourth, Tirol's strong existing emotional associations — excitement, adventure, nature — should be systematically leveraged through partnerships and influencer collaborations that extend these associations into adjacent lifestyle domains and reach audiences beyond the destination's existing visitor base. These strategic priorities are developed in detail in the Strategic Recommendations section that follows the Discussion.

Discussion

The findings of this brand audit contribute to destination branding theory and practice in several interconnected ways, and their implications extend beyond Tirol to broader questions about how mature destination brands can sustain differentiation and equity in competitive tourism environments.

Brand Awareness as a Foundational Constraint

Perhaps the most consequential finding of this study is the magnitude of the awareness gap between Tirol and Innsbruck among prospective tourists. This asymmetry illustrates a structural challenge that destination branding theory has long identified but that is rarely as clearly quantified in empirical research: when a destination's constituent cities or resorts achieve stronger independent recognition than the regional brand itself, the marketing investments of the DMO risk generating equity for individual places rather than for the overarching destination brand. As Konecnik and Gartner (2007) establish, awareness is not merely a prerequisite for consideration — it is the cognitive infrastructure upon which all other brand equity dimensions are built. Without it, even strong destination images and positive emotional associations remain poorly anchored in consumer memory and therefore fragile in the face of competitive messaging.

The Innsbruck–Tirol awareness asymmetry also carries important strategic implications that can be understood through both Keller's Customer-Based Brand Equity (CBBE) framework and destination brand architecture theory. Innsbruck's exceptionally high recognition among both current and prospective tourists suggests that it possesses substantial independent brand equity characterized by strong awareness, favorable associations, and positive emotional responses. Rather than viewing this imbalance as a weakness, Tirol can treat Innsbruck as a strategic brand asset capable of accelerating regional brand development. Consumers who already recognize and hold favorable perceptions of Innsbruck represent a reachable audience for whom stronger visual, verbal, and experiential linkages between Innsbruck and Tirol can facilitate the transfer of brand equity from the city level to the regional level.

This process, commonly referred to as equity transfer, occurs when positive associations attached to a well-known sub-brand are extended to a parent brand through deliberate co-branding and brand architecture strategies (Aaker, 1991; Keller & Swaminathan, 2019). In the destination context, co-branding enables a regional destination to leverage the recognition of constituent cities, resorts, and attractions while simultaneously reinforcing the overarching destination identity (Cai, 2002). For Tirol, this suggests that Innsbruck should not be promoted solely as an individual destination but as an integral component of the Tirol experience. Consistent use of co-branded visual identities, such as "Tirol–Innsbruck," integrated website architecture, joint promotional campaigns, and coordinated destination signage can help ensure that positive experiences associated with Innsbruck are cognitively linked to the Tirol master brand.

From a strategic perspective, this approach allows Tirol to capitalize on existing consumer knowledge structures rather than attempting to build awareness entirely from scratch. Every interaction with Innsbruck—whether through digital media, travel planning, visitation, or post-trip advocacy—becomes an opportunity to strengthen awareness and associations for Tirol as a whole. Over time, such equity transfer mechanisms can reduce the awareness gap identified in this study, strengthen regional brand coherence, and contribute to the development of stronger destination brand equity across the entire Tirol tourism ecosystem. This is precisely the logic underlying the co-branding recommendations developed from the audit findings and illustrates how destination brand equity theory can be translated into actionable destination management strategy.

The Identity–Image Gap and Its Competitive Consequences

The finding that competing destinations — particularly Graubünden and Salzburg — are more strongly associated with Tirol's own brand mantra than Tirol itself is among the most striking results of this study, and it has direct theoretical relevance to the identity–image gap literature. Hankinson (2004) argues that destination brand identity functions as a relational network of associations that must be consistently and coherently reinforced across all communications and stakeholder interactions if it is to take hold in consumer memory. The present findings suggest that this reinforcement has been insufficient: the "Heart of the Alps" concept exists as a strategic positioning statement within Tirol Werbung's brand architecture, but it has not been communicated with the frequency, consistency, or distinctiveness necessary to claim ownership of that positioning in the competitive Alpine market.

This has important implications for how DMOs should approach brand mantra management. A brand mantra is not self-executing: naming a destination the "Heart of the Alps" does not make consumers perceive it as such. The mantra must be actively and repeatedly connected to specific, tangible experiences, imagery, and emotional associations that make the positioning credible and distinctive. Where competing destinations have stronger mental availability for a given concept, reclaiming that positioning requires not merely increased communications volume but a fundamentally more integrated and strategically coherent approach to brand expression across all touchpoints.

Brand Architecture and the Fragmentation Problem

The fragmentation gap identified in this study — in which sub-destination equity fails to transfer to the Tirol master brand — reflects a broader challenge in regional destination branding documented across multiple geographic contexts (Cai, 2002; Blain, Levy, & Ritchie, 2005). Regional destination brands occupy a structurally difficult position: they must leverage the recognition and appeal of their constituent destinations to generate awareness, while simultaneously ensuring that positive associations generated at the sub-destination level accrue to the parent brand rather than remaining isolated within individual resort identities.

The present findings suggest that Tirol has thus far been more successful at the former than the latter. Its resorts are internationally recognized and generate strong visitor satisfaction, but the regional brand has not established sufficiently visible or consistent linkage mechanisms to capture the equity those resorts generate. This is a brand architecture problem as much as a communications problem, and it requires structural solutions — co-branded visual systems, consistent regional identifiers at resort entry points, integrated digital architectures — rather than simply more advertising. The strategic recommendations developed from these findings draw directly on this diagnostic insight, proposing specific co-branding mechanisms designed to make the Tirol master brand visibly present wherever its constituent destinations are experienced.

The Value of Brand Audits as Diagnostic and Strategic Tools

At a broader theoretical level, this study reinforces the value of comprehensive brand audits as diagnostic instruments within destination brand management. The findings demonstrate that routine tourism performance metrics — visitor numbers, overnight stays, and occupancy rates — would not have revealed the awareness asymmetry, mantra ownership gap, or brand architecture fragmentation identified through the audit's combination of brand inventory analysis, consumer interviews, and prospective tourist surveys. Tirol's strong arrival figures and high visitor satisfaction scores might easily suggest a healthy brand, when in fact the audit reveals structural vulnerabilities that, left unaddressed, could erode competitive positioning over time as rival destinations invest more coherently in brand differentiation.

This finding is consistent with Adidam and Shaker's (2021) argument that brand audits surface qualitative dimensions of brand health that quantitative performance indicators obscure, and with Pike's (2015) observation that destination brand equity is ultimately a consumer-mind phenomenon that cannot be inferred from supply-side metrics alone. For DMOs and policymakers, the Tirol case illustrates the practical value of embedding periodic brand audits within ongoing brand governance systems — not as one-time diagnostic exercises but as continuous learning mechanisms that keep strategic decision-making anchored in consumer reality rather than managerial assumption.

Strategic Recommendations

The following four recommendations derive directly from the empirical findings and are grounded in the destination branding theory developed throughout this study. Each recommendation is designed to address a specific gap identified in the audit while contributing to the broader objective of strengthening Tirol's destination brand equity and competitive positioning within the European Alpine tourism market.

Recommendation 1: Strengthen Competitive Positioning Through Clearly Defined Points of Parity and Points of Difference

The competitive analysis revealed substantial overlap between Tirol and its primary Alpine competitors in terms of natural scenery, outdoor recreation, and cultural heritage. While these shared attributes establish Tirol's legitimacy within the Alpine tourism category — functioning as Points of Parity (POPs) that satisfy baseline visitor expectations — they are insufficient to drive destination preference in a crowded market. The audit findings indicate that Tirol's distinctive advantages in accessibility, authenticity, affordability relative to premium competitors such as Graubünden, wellness experiences, and family-oriented tourism are not being communicated with sufficient clarity or consistency to create defensible Points of Difference (PODs).

Tirol Werbung should develop a differentiation framework that explicitly maps these PODs against each primary competitor.

- Versus Graubünden: Tirol's accessibility and inclusive pricing position it as the more welcoming Alpine experience without sacrificing authenticity.
- Versus Chamonix: Tirol's emphasis on wellness, family, and balanced recreation differentiates it from a destination defined primarily by high-adrenaline mountaineering.
- Versus Salzburg: Tirol's outdoor recreation depth and year-round activity portfolio offer a more active and immersive experience than Salzburg's predominantly cultural appeal.
- Versus South Tyrol: Tirol's unambiguous Austrian Alpine identity positions it as the archetype of the Alpine experience rather than a hybrid destination.

Communicating these distinctions consistently across digital channels, trade marketing materials, and event sponsorships would materially strengthen Tirol's competitive positioning and reduce the risk of consumer conflation with competing destinations.

Recommendation 2: Improve Brand Architecture and Visual Consistency Through Destination Co-Branding

The brand architecture fragmentation identified in the audit — in which visitors recognize Kitzbühel, Innsbruck, Ischgl, and St. Anton as individual destinations without connecting them to the Tirol master brand — represents one of the most structurally consequential gaps in Tirol's current brand equity. Addressing it requires not merely more communications but a systematic architectural solution that makes the regional brand visibly and consistently present at every sub-destination touchpoint. Specifically, Tirol Werbung should develop a unified co-branding system in which the primary Tirol tourism logo is visually paired with the name of each major resort destination — for example, Tirol–Kitzbühel, Tirol–Innsbruck, Tirol–Ischgl — displayed prominently at resort entry points, visitor centers, promotional materials, and public spaces (see Figure 4). This approach mirrors successful regional co-branding strategies documented in the destination branding literature (Cai, 2002) and would serve two strategic purposes simultaneously: transferring the positive equity accumulated at the resort level upward to the Tirol master brand, and reinforcing Tirol's salience as an overarching regional identity among visitors who may be encountering the brand for the first time.

Figure 4. Sub-Branded Logos of key resorts in Tirol

In parallel, expanding the primary tourism logo accompanied by the explicit "Heart of the Alps" mantra would strengthen cross-domain brand coherence (see Figure 5).

Figure 5. Unified Brand Logo of Tirol

Recommendation 3: Expand Awareness of the "Heart of the Alps" Positioning Through Integrated Multi-Channel Communication

The survey findings revealed that only 22% of visiting tourists and 3% of prospective tourists associate "Heart of the Alps" with Tirol — a critical deficit given that the mantra represents the cornerstone of Tirol's intended brand identity. Strengthening ownership of this positioning requires a sustained, integrated communications effort that embeds the mantra consistently across every consumer-facing channel and touchpoint.

Practically, this means ensuring that "Heart of the Alps" appears not merely in formal advertising but in social media content, event sponsorship activations, influencer campaigns, website architecture, press materials, and trade communications. The audit findings indicate that Tirol Werbung currently allocates 75% of its communications budget to digital channels — a strong structural foundation for amplifying the mantra's reach. This investment reflects a broader transformation in destination marketing practice: as Gretzel, Fesenmaier, and O'Leary (2006) observed, the shift toward digital environments has fundamentally altered how travelers discover, evaluate, and form attachments to destinations, making online presence not merely a communications channel but the primary arena in which destination brand meaning is constructed and contested. For Tirol, this means that the coherence, accessibility, and linguistic reach of its digital communications are not peripheral concerns but central determinants of whether the "Heart of the Alps" positioning achieves the consumer ownership the audit findings reveal it currently lacks. However, the predominantly German-language content of platforms such as YouTube materially limits international audience penetration, effectively concentrating the mantra's digital amplification within source markets where Tirol already enjoys relatively stronger recognition. Expanding multilingual content — particularly in English, and in the languages of emerging source markets including Italian, French, and Hindi — would significantly extend the mantra's reach among the prospective tourist segments where awareness is currently weakest and the strategic opportunity for brand equity growth is therefore greatest. Additionally, integrating "Heart of the Alps" language into Tirol's experiential branding initiatives would create physical, shareable touchpoints that organically amplify the positioning through visitor-generated content and social media dissemination.

Recommendation 4: Leverage Secondary Brand Associations Through Strategic Partnerships and Influencer Collaborations

The audit findings reveal that Tirol's strongest brand assets — excitement, adventure, nature, and authentic Alpine culture — are experiential rather than communicative: visitors feel them deeply but prospective tourists hear about them insufficiently. Bridging this gap requires extending the brand's reach into lifestyle domains and digital communities where target audiences are already engaged, through partnerships and influencer collaborations whose values and audiences align with Tirol's core brand associations.

Strategic partnerships with outdoor apparel brands, winter sports equipment manufacturers, wellness brands, and international travel media organizations would enable Tirol to extend its brand presence beyond tourism communications into the lifestyle and leisure channels through which active, health-conscious travelers increasingly discover and evaluate destinations. Co-branded campaigns with international media organizations such as National Geographic would provide high-credibility narrative exposure that reinforces Tirol's natural beauty, heritage, and sustainability positioning to global audiences. Influencer collaborations with outdoor sports enthusiasts, wellness advocates, travel bloggers, and professional ski and snowboard athletes would generate authentic, peer-mediated content that reaches prospective tourists through the social channels they trust most. Collectively, these secondary associations would facilitate brand meaning transfer — whereby the symbolic attributes of partner brands are absorbed into Tirol's own brand image — while substantially expanding the reach and emotional resonance of the destination's communications beyond its existing visitor base.

Limitations and Future Research

As with any empirical investigation, the findings of this study should be interpreted in light of several limitations. First, the study employed a mixed-methods design that relied on convenience sampling for tourist interviews conducted at selected locations within Tirol. Although Innsbruck Airport, Innsbruck Railway Station, and the Bergisel tourism area provided access to a diverse cross-section of visitors, the resulting sample may not fully represent the broader population of tourists visiting Tirol across different seasons, destinations, and travel segments. Consequently, caution should be exercised when generalizing the findings beyond the study sample.

Second, the research adopted a cross-sectional design, capturing perceptions of Tirol at a single point in time. Destination brands are dynamic and continuously influenced by marketing campaigns, media coverage, economic conditions, and changing consumer preferences. The findings therefore provide a snapshot of destination brand equity rather than an assessment of how perceptions evolve over time. Longitudinal research would provide valuable insight into the stability of brand associations and the effectiveness of branding interventions implemented by destination marketing organizations.

Third, while the study incorporated multiple data sources—including brand inventory analysis, tourist interviews, prospective tourist surveys, and stakeholder interviews—the quantitative component relied primarily on descriptive statistical analysis. The objective of the study was diagnostic rather than predictive; however, future research could employ more advanced analytical techniques such as exploratory and confirmatory factor analysis, structural equation modeling, or destination brand equity indices to further examine the relationships among awareness, image, emotional responses, loyalty, and destination choice behavior.

Fourth, the prospective tourist survey included respondents from ten countries, providing an international perspective on Tirol's brand awareness and image. Nevertheless, the sample size limited the ability to conduct detailed cross-national comparisons. Future studies could explore cultural differences in destination perceptions and evaluate whether Tirol's brand positioning resonates differently across key source markets. Such research would provide valuable guidance for market-specific branding strategies and communication campaigns.

A further limitation concerns the stakeholder interview component. Although tourism operators and travel intermediaries provided valuable industry perspectives, their views may not fully reflect the perspectives of other important stakeholder groups, including residents, hospitality employees, local businesses, and policymakers. Future destination branding research would benefit from incorporating a broader stakeholder perspective to better understand how destination identity is constructed, communicated, and experienced across the tourism ecosystem.

Finally, this study focused exclusively on Tirol as a single destination case. While case study research provides rich contextual insight, future comparative studies examining multiple Alpine destinations—such as Salzburg, South Tyrol, Graubünden, and Chamonix—would enable researchers to identify broader patterns of destination brand equity development and competitive positioning. Comparative research could also provide deeper insight into how destinations establish ownership of brand mantras, manage brand architecture, and differentiate themselves within highly competitive tourism regions.

Despite these limitations, the study demonstrates the value of applying a comprehensive brand audit framework to destination branding research. Future research that extends this approach across destinations, stakeholder groups, and time periods can further enhance understanding of how destination brand equity is created, maintained, and strengthened in increasingly competitive tourism environments.

Conclusion

This study applied a comprehensive destination brand audit to evaluate Tirol's brand equity and competitive positioning within the European Alpine tourism market. Guided by destination branding theory, Konecnik and Gartner's (2007) multi-dimensional brand equity framework, and Keller's Customer-Based Brand Equity model, the study examined the alignment between Tirol's intended brand identity and consumer perceptions through a mixed-methods research design integrating brand inventory analysis, tourist interviews, prospective tourist surveys, and stakeholder consultations.

The findings indicate that Tirol possesses genuine and substantial destination brand assets. Its associations with nature, alpine sports and adventure, and Tyrolean tradition and culture are strong, consistent, and emotionally resonant — generating the kind of excitement, fun, and adventure responses that, according to the CBBE framework, form the affective foundation of deep brand loyalty. These are not trivial strengths: they represent decades of accumulated destination equity derived from high-quality visitor experiences, world-class tourism infrastructure, and a distinctive cultural identity that competitors cannot easily replicate.

Yet the audit also reveals that these assets are incompletely leveraged. Three structural gaps — limited brand awareness among prospective tourists, weak marketplace ownership of the "Heart of the Alps" mantra, and fragmented brand architecture between individual resorts and the regional parent brand — collectively constrain Tirol's ability to translate exceptional visitor experiences into proportionate destination brand equity. The most telling diagnostic finding is the awareness asymmetry between Tirol and Innsbruck: a gap that reveals not a failure of destination quality but a failure of brand architecture and communications coherence. Tirol is generating tourism but not fully generating brand equity — and in an increasingly competitive Alpine marketplace, that distinction matters enormously for long-term competitiveness.

The four strategic recommendations developed from these findings — clarifying Points of Parity and Points of Difference, establishing unified co-branding architecture, intensifying multi-channel promotion of the "Heart of the Alps" positioning, and leveraging secondary associations through strategic partnerships and influencer collaborations — provide an integrated and evidence-based framework for closing these gaps. Critically, each recommendation is anchored in the empirical findings of the audit rather than in generic best practice, ensuring that strategic investment is directed toward the specific vulnerabilities the data reveal.

At a broader scholarly level, this study contributes to destination branding research in three ways. First, it demonstrates the application of a systematic two-stage brand audit framework — combining brand inventory and brand exploratory analysis — within a mature Alpine destination context, responding to calls for more holistic, mixed-methods approaches to evaluating destination brand equity (Hankinson, 2004; Hanna & Rowley, 2015). Second, it provides empirical evidence of the identity–image gap dynamic in a real destination context, illustrating how intended brand positioning can fail to achieve consumer ownership even when the underlying destination experience is strong and broadly consistent with the intended identity. Third, it illustrates how brand audit findings can be translated into theoretically grounded strategic recommendations — demonstrating that brand audits are not merely evaluative exercises but generative strategic processes through which destinations can move from assumption to evidence and from fragmentation to coherence.

For destination marketing organizations and policymakers, the Tirol case offers a transferable lesson: strong arrival figures and high visitor satisfaction are necessary but not sufficient indicators of destination brand health. Structural brand equity — the awareness, associations, and loyalty that sustain competitive advantage across market cycles — requires deliberate, coherent, and continuously evaluated brand management. Periodic brand audits, embedded within ongoing governance systems rather than conducted as isolated exercises, provide the diagnostic foundation that makes such management possible.

As the Alpine tourism landscape grows more competitive and travelers grow more discerning in their destination choices, the destinations that endure will be those that not only deliver exceptional experiences but ensure that those experiences are reliably connected to a clear, distinctive, and emotionally resonant brand identity. The path forward for Tirol is not to reinvent itself but to more fully and coherently claim the exceptional destination it already is — to ensure that the heart of the Alps beats loudly enough for the world to hear.

Acknowledgments

This research was made possible through the support of the *Fulbright Scholar Program*, sponsored by the U.S. Department of State's Bureau of Educational and Cultural Affairs. The author also thanks Tirol Werbung GmbH for providing access to official reports, website content, and communication materials that informed the brand audit of Tirol.

References

Aaker, D. A. (1991). *Managing brand equity: Capitalizing on the value of a brand name*. Free Press.

Adidam, P. T., & Shaker, M. (2021). Brand audits: A critical tool for strategic brand management. *Journal of Brand Strategy, 10*(4), 367–381.

Baloglu, S., & McCleary, K. W. (1999). A model of destination image formation. *Annals of Tourism Research, 26*(4), 868–897.

Blain, C., Levy, S. E., & Ritchie, J. R. B. (2005). Destination branding: Insights and practices from destination management organizations. *Journal of Travel Research, 43*(4), 328–338.

Boo, S., Busser, J., & Baloglu, S. (2009). A model of customer-based brand equity and its application to multiple destinations. *Tourism Management, 30*(2), 219–231.

Buhalis, D. (2000). Marketing the competitive destination of the future. *Tourism Management, 21*(1), 97–116.

Cai, L. A. (2002). Cooperative branding for rural destinations. *Annals of Tourism Research, 29*(3), 720–742.

Gretzel, U., Fesenmaier, D. R., & O'Leary, J. T. (2006). The transformation of consumer behaviour. In D. Buhalis & C. Costa (Eds.), *Tourism business frontiers* (pp. 9–18). Butterworth-Heinemann.

Hankinson, G. (2004). Relational network brands: Towards a conceptual model of place brands. *Journal of Vacation Marketing, 10*(2), 109–121.

Hanna, S., & Rowley, J. (2015). Towards a model of the place brand web. *Tourism Management, 48*, 100–112.

Hudson, S., Roth, M. S., Madden, T. J., & Hudson, R. (2015). The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees. *Tourism Management, 47*, 68–76.

Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing, 57*(1), 1–22.

Keller, K. L., & Swaminathan, V. (2019). *Strategic brand management: Building, measuring, and managing brand equity* (5th ed.). Pearson.

Konecnik, M., & Gartner, W. C. (2007). Customer-based brand equity for a destination. *Annals of Tourism Research, 34*(2), 400–421.

MCI Management Center Innsbruck. (2025, March 10). *Key facts im Tiroler Tourismus*. FACT.tirol. https://www.fact.tirol/statistik/key-facts-im-tiroler-tourismus/

Mariani, M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional DMO pages. *Tourism Management, 54*, 321–343.

Morgan, N., Pritchard, A., & Pride, R. (Eds.). (2011). *Destination brands: Managing place reputation* (3rd ed.). Butterworth-Heinemann.

Pike, S. (2015). *Destination marketing: Essentials*. Routledge.

Pike, S., & Page, S. J. (2014). Destination marketing organizations and destination marketing: A narrative analysis of the literature. *Tourism Management, 41*, 202–227.

Ruzzier, M., & de Chernatony, L. (2013). Developing and applying a place brand identity model: The case of Slovenia. *Journal of Business Research, 66*(1), 45–52.

Tirol Werbung GmbH. (2025). *History of the Tirol brand*. https://www.tirol.at

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58*, 51–65.

Persistence or Pivot: A Mixed-Methods Analysis of Linguistic Markers and Vocational Identity in Introductory Field Experiences

by

Lindon Ratliff, PhD

Professor of Secondary Education

Mississippi State University

This study introduces the concept of the early field placement as a ”vocational stress test”, a critical diagnostic period where pre-service teachers reconcile their idealized professional identity with the authentic realities of the secondary classroom during an initial 15-hour placement in an introductory secondary education course. Utilizing a mixed methods approach, the research identifies critical linguistic and behavioral markers that differentiate students who persist in the major from those who pivot to other disciplines. While initial reality shock and environmental anxiety are universal in the first month, the second month serves as a pivotal period where behavioral patterns emerge; students who stay in the major are characterized by a transition from passive observation to active classroom participation. Qualitative content analysis reveals that these stayers undergo a significant shift from clinical detachment to relational ownership marked by the adoption of possessive language regarding their students whereas leavers maintain a stationary observer only status throughout the duration of the placement. These findings suggest that early attrition is often a successful diagnostic outcome of the field experience rather than a programmatic failure, and the study recommends utilizing midpoint journal audits as a proactive tool for teacher preparation programs to identify and intervene with disengaged candidates.

Key words: Teacher retention, field, secondary education, pre-service teachers

Introduction

The looming crisis of teacher attrition is often framed through the lens of veteran burnout or the departure of early-career professionals within their first five years of service. However, a critical yet under-examined segment of this attrition occurs much earlier, during the initial socialization phase of teacher preparation programs. In this study, we propose viewing this introductory phase as a vocational stress test, where pre-service teachers move from the theoretical comfort of the university classroom into the complex, unpredictable environment of the secondary school. For many candidates, this first exposure triggers what Veenman (1984) famously termed reality shock, forcing a reconciliation between their idealized perceptions of teaching and the logistical and emotional demands of the contemporary classroom.

The transition from student to practitioner is not merely a matter of accruing clinical hours; it is a process of identity construction. During this period, candidates navigate what we term a shift from being a clinical reporter of classroom activities to becoming an active participant in the instructional community. This research seeks to understand the markers of this transition by analyzing the longitudinal field journals of 54 secondary education students. By tracking the evolution of their linguistic choices and behavioral engagement across a semester, this study identifies the pivotal moments where a student either finds what we conceptualize as a relational hook that leads to persistence or experiences a vocational pivot toward an alternative discipline.

Understanding these early indicators is essential for teacher education faculty and field coordinators. If the factors that predict major retention can be identified within the first 10 hours of observation, programs can move toward a more proactive model of candidate support. This study suggests that early attrition should not be viewed as a programmatic failure, but rather as a successful diagnostic function of the field experience. By examining the divergence between those who stay and those who leave, we can better understand how to foster professional resilience and ensure that the pipeline to the teaching profession is populated by candidates who have successfully integrated into the relational reality of the classroom.

Purpose of the Study

The purpose of this study is to examine the longitudinal development of vocational identity among pre-service teachers by identifying specific linguistic and behavioral markers that correlate with enrollment persistence. Specifically, the research seeks to analyze the transition from initial reality shock to professional divergence, investigating how the shift from clinical detachment to relational ownership, operationalized through the adoption of possessive language and active instructional engagement, serves as a leading indicator of major retention. By quantifying the relationship between early field participation and subsequent major changes, this study aims to provide teacher education programs with a proposed diagnostic framework for identifying at risk candidates at a critical midpoint, ultimately reframing early phase attrition as a necessary mechanism for ensuring professional fit and long-term career resilience.

Research Questions

3. To what extent does the level of active engagement during introductory field observations correlate with a candidate's decision to persist in or pivot from the secondary education major?
4. How do the linguistic markers in pre-service teacher journals evolve from initial reality shock to vocational resolution, and how do these shifts differ between stayers and leavers?
5. In what ways do critical incidents and the observation of mentor behaviors act as a mirror for candidates, influencing their perceptions of professional fit and career longevity?
6. To what degree can the transition from clinical detachment to relational ownership be identified as a primary diagnostic indicator of vocational commitment by the midpoint of the initial field experience?

Review of Relevant Literature

Theoretical Foundations of Preservice Teacher Socialization

The transition into the teaching profession is a multidimensional process of socialization that commences long before a candidate assumes formal instructional responsibilities. Socialization, in this context, is defined as the process by which an individual acquires the social knowledge and skills necessary to assume a professional role (Danziger, 1971). Recent scholarship emphasizes that this process is increasingly mediated by the pre-professional identity candidates bring to their teacher education programs, which functions as a cognitive filter for their initial classroom observations (Izadinia, 2018). Within teacher education, this trajectory is frequently analyzed through Social Cognitive Career Theory, which posits that career interest and persistence are governed by a reciprocal relationship between self-efficacy beliefs, outcome expectations, and personal goals (Lent, Brown, & Hackett, 1994).

Central to this socialization is the concept of Legitimate Peripheral Participation (Lave & Wenger, 1991). This theory suggests that newcomers to a professional community begin on the periphery (observing and performing low-risk tasks) before gradually migrating toward more central, complex roles. In the context of the current study, the fifteen-hour observation period represents this critical peripheral stage. Modern research indicates that the quality of these initial interactions is a more potent predictor of long-term retention than the sheer volume of clinical hours (Goldhaber et al., 2022). Successful socialization is therefore marked by a candidate's ability to move beyond a stationary, observer-only status and begin identifying with the professional community of practice (Wenger, 1998).

Furthermore, the construction of a professional identity is a linguistic and psychological project. As candidates engage with the classroom environment, they negotiate a new persona that bridges their previous identity as a student with their emerging identity as an educator (Britzman, 2003). Recent longitudinal studies identify the early adoption of professional language (specifically the shift from external descriptions of students to internal, relational narratives) as a key indicator of vocational resilience in high-pressure secondary environments (Nguyen & Tran, 2023). This negotiation is often reflected in the narrative shifts within reflective journals; when a candidate successfully socializes, their reflections move from detached descriptions of classroom mechanics to internalized accounts of student interaction (Sachs, 2005).

Reality Shock and the Transition to Relational Ownership

The initial entry into the secondary school environment is frequently characterized by what Veenman (1984) termed reality shock; the collapse of idealistic expectations when confronted with the logistical and behavioral complexities of the classroom. Recent research indicates that this shock is often intensified by environmental anxiety, where candidates focus on the physical and bureaucratic infrastructure of the school rather than pedagogical execution. This state of heightened awareness regarding school safety protocols and administrative rigidity can create a state of instructional blindness during the first five hours of observation.

This anxiety serves as a primary barrier to the socialization process; when candidates are overwhelmed by the environmental hardening of modern schools, they often remain in a state of clinical detachment (Lortie, 1975; Zeichner & Tabachnick, 1981). Addressing this reality shock is essential, as the inability to move past environmental anxiety prevents candidates from engaging with the relational aspects of the classroom. Consequently, the introductory field experience serves as a significant site of emotional negotiation where candidates must reconcile their preconceived notions of schooling with the contemporary secondary reality (Zembylas, 2003).

Building upon the framework of situated learning, the transition toward the center of the professional community requires more than passive presence; it requires the establishment of a relational hook. While a student on the periphery remains a clinical observer of classroom mechanics, a student moving toward the center begins to find a specific emotional or professional connection to the students being served. This movement is a decisive factor in vocational persistence.

Contemporary research suggests that candidates who do not move beyond the role of a stationary observer by the midpoint of their placement are at a significantly higher risk of vocational exit (Goldhaber et al., 2020; Ingersoll, 2001). This lack of engagement often stems from an inability to reconcile an emerging professional identity with the complex social dynamics of the classroom (Hong, 2010). Without a relational hook, field experience remains an abstract academic exercise rather than a transformative milestone (Palmer, 1998). Furthermore, the adoption of an instructional identity is heavily dependent on the willingness of the candidate to assume micro-responsibilities (such as circulating during independent work) which serves to anchor them within the community of practice (Lave & Wenger, 1991).

Linguistic Markers and the Construction of Professional Identity

Professional identity is a semiotic project where the language utilized by a candidate reflects their evolving relationship with the profession. According to Gee (2014), identity is enacted through specific discourses that signal membership within a professional group. For the pre-service teacher, the reflective journal serves as a primary site for this discursive work. Narrative choices provide a window into whether candidates view the classroom as outside critics or emerging members of the instructional team.

A critical indicator of this transition is the shift from clinical detachment to relational ownership. Early in the field experience, candidates often utilize objective language, referring to students as "the learners" or "the class." However, as they establish a relational hook, their language typically shifts to possessive pronouns such as "my students" or "our classroom." These minor linguistic pivots are not merely stylistic; they indicate an internal shift in professional self-efficacy and a burgeoning sense of accountability for student success (Nguyen & Tran, 2023). Such "ownership markers" are highly correlated with vocational resilience (Izadinia, 2018), while those who maintain a third-person perspective, focusing exclusively on infrastructure, are more likely to experience a vocational pivot.

The Mentor Teacher as a Proximal Mirror

The mentor teacher functions as a proximal mirror through which the pre-service teacher projects their own future professional identity. According to Bandura (1997), vicarious experience is a primary source of self-efficacy. By observing a veteran practitioner, the candidate performs a mental self-substitution, assessing whether they possess the skills to navigate similar challenges and whether the educator's lifestyle is a desirable long-term outcome. In this context, the mentor becomes the primary data point for the candidate's vocational calculations. Simons & Zhang (2024) stated the observer is an active agent performing a longitudinal calculation of career longevity. This mirroring process is particularly critical when candidates witness high-stress incidents or teacher exhaustion. The emotional state of the mentor acts as a vocational filter; while seeing a master teacher manage a crisis provides a sense of possibility, witnessing chronic fatigue can trigger vicarious burnout (Goldhaber et al., 2022). High-aptitude candidates may engage in a logical cost-benefit analysis, determining that the professional return on investment is insufficient (Nguyen & Tran, 2023). Therefore, the initial field experience is as much an observation of the professional lifestyle as it is of instructional technique.

Early Attrition as a Mechanism for Vocational Alignment

In the broader discourse of teacher education, student attrition is frequently categorized as a programmatic failure. However, contemporary scholarship challenges this deficit perspective, suggesting that early departures may represent a successful diagnostic function of well-designed clinical experiences. By providing an authentic stress test prior to high-stakes environments, introductory field placements allow for vocational alignment, ensuring that those who proceed possess the necessary resilience (Goldhaber et al., 2022).

This perspective shifts the focus from simple retention metrics to the quality of the vocational fit. When candidates exit the major after initial exposure, they are performing a sophisticated self-assessment of their interpersonal and emotional compatibility with the profession (Fowler et al., 2023). Reframing early attrition as a diagnostic success allows faculty to move beyond GPA as the sole metric of readiness, instead monitoring the behavioral and linguistic markers that signal successful integration. Ultimately, the goal of introductory field work is not to retain every student, but to facilitate an informed decision-making process that aligns individual career goals with the rigorous demands of the modern educational landscape.

Methodology

This study employed a mixed methods longitudinal design to investigate the relationship between early field experiences and major retention among a cohort of secondary education candidates (N=54). Data collection spanned a three year period and consisted of two primary streams: qualitative field journals (1 semester) and quantitative enrollment records (over a three year period).

Participant Selection and Context

The participants (n=54) were undergraduate students enrolled across three consecutive academic years in an introductory secondary education course at a public, research-intensive university in the Southeastern United States. Demographically, the cohort was composed of 68.5% female (n=37) and 31.5% male (n=17) students, with a mean age of 19.8 years (SD=1.2, range 18-23). In terms of racial and ethnic background, participants identified as White/Caucasian (74.1%, n=40), Black/African American (18.5%, n=10), Hispanic/Latino (3.7%, n=2), and Multiracial/Other (3.7%, n=2). Academically, the majority of candidates held sophomore standing (61.1%, n=33) or freshman standing (25.9%, n=14), with a smaller subset of juniors (13.0%, n=7). Participants represented diverse secondary discipline tracks, including English Language Arts (31.5%, n=17), Social Studies (29.6%, n=16), Mathematics (20.4%, n=11), and Natural Sciences (18.5%, n=10). Field placements occurred across diverse secondary settings, ranging from rural Title I schools to larger suburban districts. This variety ensured that the findings regarding reality shock and environmental perception were not limited to a single socioeconomic context.

Data Collection Instruments

The primary data source consisted of four sequential reflective folders submitted by each participant throughout the semester. These included:

h. Folder 1: Initial observations focused on the physical environment and safety protocols (Hours 1 through 5).

i. Folder 2: Midpoint reflections detailing instructional engagement and relational dynamics (Hours 6 through 10).

j. Folder 3: Advanced observations focusing on professional identity and classroom management (Hours 11 through 15).

k. Folder 4: A final synthesis letter providing advice to future candidates and a definitive statement of vocational intent**.**

Analytical Procedures

Qualitative data analysis followed a multi-phase coding process guided by inductive and deductive content analysis techniques. In Phase 1, two researchers independently conducted open coding on the full corpus of longitudinal field journals ($N = 54$) to identify initial codes related to environmental perceptions, anxiety, mentor interactions, and student engagement. In Phase 2, axial coding was applied to collapse initial codes into primary categories, specifically tracking the longitudinal trajectory of linguistic choices (e.g., third-person clinical descriptions vs. possessive pronouns) and behavioral involvement across the four sequential folders.

Trustworthiness and Inter-Coder Agreement

To establish qualitative trustworthiness and analytical rigor, several procedures were implemented:

e. **Inter-Coder Reliability:** To evaluate coding consistency, a randomly selected 20% sample of the transcript corpus ($n = 11$) was independently double-coded by a second researcher. Inter-coder agreement was calculated using Cohen's kappa (κ), yielding an initial agreement rate of $\kappa = .84$, which indicates strong inter-coder reliability prior to resolving minor discrepancies through consensus meetings.

f. **Member Checking and Triangulation:** Qualitative interpretations were triangulated across four sequential journal submissions per participant (Folders 1–4) and cross-referenced with mentor teacher feedback forms.

g. **Audit Trail:** A detailed codebook containing operational definitions, inclusion/exclusion rules, and representative text snippets was maintained throughout the analysis to ensure complete analytical auditability.

Mixed-Methods Integration

This study utilized a convergent mixed-methods design with quantitizing procedures. Qualitative findings were integrated with quantitative institutional enrollment records through a systematic data-transformation strategy. Specifically, core qualitative categories identified during axial coding (e.g., the adoption of possessive language by Hour 10 and the occurrence of ≥3 active classroom engagements) were transformed into binary dummy variables (0 = absent, 1 = present). These quantitized variables were then merged with subsequent institutional enrollment data (0 = pivoted / exited major, 1 = persisted in major) to perform point-biserial correlations and cross-tabulation analyses. This integration directly connected the longitudinal qualitative narrative evolution to hard institutional retention outcomes.

Results

Descriptive Statistics and Overview

Prior to conducting inferential analyses, descriptive statistics were calculated for all quantitized qualitative variables and institutional enrollment outcomes across the full cohort (n=54). Overall, 81.5% of the cohort (n=44) persisted in the secondary education major into the subsequent academic year, while 18.5% (n=10) executed a vocational pivot to an alternative discipline.

By the 10-hour midpoint (Folder 2/3), 66.7% of all participants (n=36) demonstrated the adoption of possessive language (e.g., "my students," "our classroom"). Among those who ultimately persisted in the major (n=44), 81.8% (n=36) adopted possessive language, whereas

0.0% (n=0) of those who pivoted (n=10) adopted possessive language. Furthermore, active behavioral engagement (≥3 documented instances of active participation) was observed in 70.4% of the total cohort (n=38). Active engagement was present in 86.4% of stayers (n=38) compared to 0.0% of leavers (n=0).

Table 1 summarizes the descriptive statistics and distribution of key quantitative variables by enrollment persistence status.

Table 1

Variable	**Total Cohort (n=54)**	**Stayers (n=44)**	**Leavers (n=10)**
Possessive Language Adoption (by Hr 10)	66.7% (n=36)	81.8% (n=36)	0.0% (n=0)
Active Engagement ($\geq 3\ instances$)	70.4% (n=38)	86.4% (n=38)	0.0% (n=0)
Mentor Exhaustion	11.1% (n=6)	0.0% (n=0)	60% (n=6)

The September Baseline: Universal Reality Shock (Addressing RQ2)

Data from the first five hours of observation (Folder 1) addressed RQ2 by establishing a universal baseline of reality shock and environmental anxiety across 100% of the cohort. Qualitative coding indicated that 92% of participants focused exclusively on the physical and bureaucratic infrastructure of the school. Frequent descriptors included "metal detectors," "security protocols," and "locked doors." During this phase, 100% of candidates utilized detached, third-person language when referring to students, with the most frequent identifiers being "the learners" or "the kids." These findings suggest that reality shock and environmental anxiety function as a normative baseline for introductory field experiences.

Sample Student Quote:

"I felt overwhelmed immediately walking past the security desks and locked doors; it felt more like a fortress than a classroom." (Participant 12)

The October Pivot: Linguistic and Behavioral Divergence (Addressing RQ1 & RQ2)

In addressing RQ1 (active engagement) and RQ2 (linguistic evolution), a primary finding of this study is the emergence of a professional divergence at the ten-hour midpoint (Folder 2 and Folder 3). Quantitative analysis of enrollment data revealed that 18% of the cohort (n=10) initiated a change of major or "pivot" following the conclusion of the field experience. When comparing the journals of those who persisted (stayers) with those who exited (leavers), significant linguistic and behavioral variances were identified:

f. **Stayers:** Participants who persisted in the major demonstrated a 74% increase in the use of possessive pronouns (e.g., "my students," "our lesson") between Hour 5 and Hour 15. Furthermore, these candidates documented a minimum of three instances of active participation, such as circulating the room or assisting with small-group instruction.

g. **Leavers:** Participants who subsequently exited the major maintained a 0% shift in possessive language, continuing to utilize clinical, detached descriptors through the final journal entry. These candidates also reported significantly fewer instances of active engagement, remaining in a stationary, observer-only status.

Sample student quotes:

e. **Stayer:** *"Circulating today, I helped* ***my students*** *work through the algebra problems, and it finally felt like* ***our classroom****."* (Participant 34)

f. **Leaver:** *"I sat in the back observing* ***the kids*** *again today. I still feel like a visitor watching someone else's job."* (Participant 07)

The Mirror Effect and Vicarious Burnout (Addressing RQ3)

To answer RQ3 regarding the influence of mentor behaviors, qualitative synthesis of Folder 4 revealed that the mirror effect served as a primary diagnostic tool for candidates who executed a vocational pivot. Within the group that pivoted, 60% cited the exhaustion or stress of the mentor teacher as a decisive factor in their departure. These candidates performed a mental self-substitution, concluding that the observed lifestyle of the veteran educator did not align with their long-term quality-of-life expectations. In contrast, stayers framed mentor challenges as problems to be solved rather than reasons to exit, indicating a higher degree of relational anchoring and vocational resilience.

Sample student quote: *"My mentor teacher is exhausted every single day. Seeing how consumed she is makes me realize I don't want this lifestyle."* (Participant 19)

Statistical Correlation of Diagnostic Markers (Addressing RQ4)

To evaluate the relationship between midpoint diagnostic markers and subsequent enrollment retention, inferential statistical tests were conducted on the quantitized dataset ($N = 54$). Because both possessive language adoption at Hour 10 ($0 = absent$, $1 = present$) and enrollment persistence ($0 = pivoted$, $1 = persisted$) were measured as binary categorical variables, a point-biserial correlation (r_{pb}) and Chi-square test of independence (χ^2) were performed. The point-biserial correlation revealed a strong, statistically significant positive association between the adoption of possessive language by Hour 10 and enrollment persistence into the subsequent academic year ($r_{pb} =.68$, $p <.001$, 95% CI $[.51,.80]$).

A Fisher’s Exact Test of Independence further confirmed that candidates who adopted possessive language by the 10-hour mark were significantly more likely to persist in the major than those who did not ($p <.001$, $Cramér's\ V =.68$). Similarly, a Fisher's Exact Test indicated a significant association between active behavioral engagement (≥3 active instances) and enrollment persistence ($p < .001$, $Cramér's\ V = .73$). These statistical properties confirm that early linguistic and behavioral markers carry high predictive validity for candidate retention, functioning as robust diagnostic indicators rather than mere qualitative artifacts.

Discussion

The results of this study suggest that the introductory field experience serves as a critical vocational filter, moving beyond the traditional role of a pedagogical observation period to function as a primary diagnostic site for professional fit. By analyzing the findings in direct relation to the research questions and the established literature, several implications for teacher education programs emerge. Because this study utilized an observational design, these findings demonstrate strong empirical associations rather than direct causal or predictive relationships. Early linguistic shifts and active participation should be interpreted as observable behavioral indicators that co-occur with major persistence, rather than deterministic mechanisms that guarantee candidate retention.

Active Engagement and the Transition from the Periphery (RQ1)

The data directly address RQ1, which examined the correlation between active engagement and persistence. While all candidates began as passive observers, those who persisted (stayers) documented a minimum of three instances of active instructional participation by the ten-hour mark. This behavioral shift represents the practical application of Legitimate Peripheral Participation (Lave & Wenger, 1991). As candidates moved from the back of the room to the instructional floor, they transitioned from being "observers of" to "participants in" the community of practice. Conversely, candidates who resisted this movement remained in a state of clinical detachment, failing to anchor themselves within the relational fabric of the school (Sachs, 2005).

Linguistic Evolution and the Semiotic Project of Identity (RQ2)

In response to RQ2, the study successfully mapped the linguistic evolution of the cohort. The transition from the "reality shock" of Folder 1 to the vocational resolution in Folder 4 was marked by a significant divergence in pronoun usage. Stayers demonstrated a 74% increase in the mean frequency of possessive pronouns (e.g., "my students," "our classroom") per 1,000 words from Folder 1 (Hours 1–5) to Folder 3 (Hours 11–15). This metric was calculated by normalizing possessive pronoun counts per participant journal submission to account for narrative length, calculating the subgroup mean at Baseline (Folder 1) versus Midpoint/Late placement (Folders 2 and 3), and determining the relative percentage change. In contrast, leavers exhibited a 0% shift in normalized possessive language usage across the same time frame. This finding validates the work of Nguyen and Tran (2023), confirming that professional identity construction is a linguistic project. The shift from "the students" to "my students" is not merely stylistic; it is a semiotic signal that the candidate is negotiating a new persona that bridges their previous identity as a student with their emerging identity as an educator (Britzman, 2003).

The Proximal Mirror and Vicarious Burnout (RQ3)

The qualitative synthesis of Folder 4 addressed RQ3 by illustrating how candidates project themselves into the reality of their mentors. For candidates who ultimately pivoted despite strong initial academic performance, the mentor served as a proximal mirror that reflected vicarious burnout rather than professional possibility (Bandura, 1997). These candidates performed a mental self-substitution, concluding that the observed lifestyle of the veteran educator, characterized by high stress and exhaustion, did not align with their long-term quality of life expectations. This alignment, or lack thereof, is a key component of Social Cognitive Career Theory, where outcome expectations and personal goals govern persistence

(Lent, Brown, & Hackett, 1994). As identified by Simmons and Zhang (2024), the observer is an active agent performing a longitudinal calculation of career longevity based on these reflections.

Relational Ownership as a Diagnostic Framework (RQ4)

Finally, the statistical correlation (r = .68) addresses RQ4, confirming that relational ownership is a primary diagnostic indicator of vocational commitment. By the ten-hour midpoint, the presence of ownership markers co-occurred consistently with subsequent enrollment persistence. This provides empirical support for reframing early attrition as a diagnostic success (Goldhaber et al., 2022). Rather than a failure, the "October Pivot" serves as a crucial vocational stress test. It offers an authentic look at the job, allowing candidates to test their emotional readiness and ensuring the secondary teaching pipeline is built on genuine resilience (Fowler et al., 2023)

Conclusion

The results of this longitudinal study demonstrate that the introductory field experience is not merely a logistical prerequisite but a sophisticated vocational stress test. While the initial five hours are characterized by a universal state of reality shock (Veenman, 1984), the subsequent ten hours reveal a profound professional divergence. This divergence is quantitatively mapped through the adoption of possessive linguistic markers and qualitatively signaled through the establishment of a relational hook. The finding that a shift in pronoun usage (from the clinical "the students" to the relational "my students") is statistically significantly correlated with retention (r = .68) suggests that identity construction is a visible, measurable project occurring well before the residency year.

Furthermore, the data regarding the mirror effect indicates that candidates who executed a vocational pivot are active agents in their own socialization, utilizing the mentor teacher as a proximal mirror to calculate future career longevity (Bandura, 1997). When this mirror reflects chronic exhaustion rather than professional agency, the resulting vicarious burnout triggers a logical vocational pivot. Viewed this way, these early departures are not a sign of program failure, but evidence of the field experience working exactly as it should. By serving a critical diagnostic function, this phase screens for the deep relational resilience and vocational alignment necessary to survive and thrive in today's secondary classrooms.

Recommendations for Teacher Education Programs

Based on the findings of this study, the following recommendations are proposed for secondary education departments at research-intensive institutions:

f. **Implement Linguistic Auditing of Reflective Journals:**

Departments should move beyond holistic grading of field journals and implement systematic linguistic monitoring. By identifying the presence or absence of ownership markers by the ten-hour midpoint, faculty can recognize at-risk candidates early. Candidates who remain anchored in detached, third-person reporting should receive targeted advising to determine if the lack of a relational hook stems from environmental anxiety or a fundamental lack of vocational fit.

g. **Mandate "Micro-Participation" Milestones:**

To accelerate the transition from the periphery to the center of the community of practice (Wenger, 1998), programs should require specific active-engagement tasks during the first ten hours. These may include assisting with small-group transitions, circulating during independent work, or leading brief instructional segments. These micro-engagements catalyze the formation of the relational hook, moving the candidate from a passive observer to an active participant in student success.

h. **Optimize Mentor Selection for Vocational Modeling:**

Placement coordinators should prioritize mentors who demonstrate not only pedagogical excellence but also professional sustainability. Given that candidates use mentors as a proxy for their own future quality of life, it is essential that the "proximal mirror" reflects a career path that is emotionally and professionally durable. Programs should also incorporate structured debriefing sessions where candidates can process the "mirror effect" with faculty to differentiate between the challenges of a specific site and the broader realities of the profession.

i. **Institutionalize "Off-Ramp" Advising Protocols:**

Universities should formalize a process for validating the decision to pivot from the education major. By framing early attrition as a successful diagnostic outcome of the field experience, departments can protect the integrity of the professional pipeline while supporting the student's transition to an alternative discipline. This reduces the systemic costs associated with late-stage attrition and ensures that the teaching force is populated by candidates who have successfully integrated into the relational reality of the secondary school.

References

Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.

Britzman, D. P. (2003). Practice makes practice: A critical study of learning to teach. State University of New York Press.

Danziger, K. (1971). Socialization. Penguin.

Fowler, S., et al. (2023). Beyond seat time: Measuring the impact of active participation in early clinical experiences. Journal of Teacher Education, 74(3), 215-229.

Gee, J. P. (2014). An introduction to discourse analysis: Theory and method. Routledge.

Goldhaber, D., Krieg, J., & Theobald, R. (2022). Exploring the investiture of student teaching: How teacher candidates' perceptions of their student teaching experience relate to their career outcomes. Journal of Teacher Education, 73(4), 366-381.

Goldhaber, D., Theobald, R., & Tien, C. (2020). Why we need a better understanding of the clinical component of teacher preparation. *Phi Delta Kappan*, *102*(3), 28–31. https://doi.org/10.1177/0031721720970698

Hong, J. Y. (2010). Pre-service and beginning teachers' professional identity and its relation to dropping out of the profession. *Teaching and Teacher Education*, *26*(8), 1530–1543. https://doi.org/10.1016/j.tate.2010.06.003

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499–534. https://doi.org/10.3102/00028312038003499

Izadinia, M. (2018). Pre-service teachers' professional identity and its relation to self-efficacy and well-being. Journal of Education for Teaching, 44(5), 635-648.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. Journal of Vocational Behavior, 45(1), 79–122.

Lortie, D. C. (1975). *Schoolteacher: A Sociological Study*. University of Chicago Press.

Nguyen, M. H., & Tran, H. (2023). Professional identity development of pre-service teachers: A longitudinal study through the lens of identity-in-practice. Teaching and Teacher Education, 122, 103945.

Palmer, P. J. (1998). *The courage to teach: Exploring the inner landscape of a teacher's life*. Jossey-Bass.

Sachs, J. (2005). The activist teaching profession. Open University Press.

Simmons, R., & Zhang, L. (2024). The agency of the observer: Identity construction in introductory field placements. Teaching and Teacher Education, 138, 104412.

Wenger, E. (1998). Communities of practice: Learning, meaning, and identity. Cambridge University Press.

Zembylas, M. (2003). Emotions and teacher identity: A poststructural perspective. *Teachers and Teaching*, *9*(3), 213–238. https://doi.org/10.1080/13540600309378

Cooking Up Connections with Rachael Ray: How the Talk Show Host Engages Her Audience Through Parasocial Interaction

Janet Taylor Birkey, *Eastern New Mexico University*
Darrell L. Roe, *Eastern New Mexico University*
Leonard Bailey, *Eastern New Mexico University*

Abstract

After seventeen seasons and over 2000 first-run episodes, *The Rachael Ray Show* continues to build connections even in second-run syndication through the parasocial interaction effect initiated by the host, cooking demonstrator and interviewer, Rachael Ray. Parasocial interaction (PSI) theory explains that viewers will develop feelings of genuine social interaction and begin to believe they have a relationship with television show hosts, particularly with those who intentionally engage their viewers in a personable manner. Six distinct categories of parasocial engagement initiated by Ray were discovered through observational data collected from ten randomly sampled episodes from two seasons of *The Rachael Ray Show*. These were articulated as 1) making eye contact, 2) responding thoughtfully, 3) verbally building rapport, 4) making herself appreciable, 5) follow-along descriptions, and 6) show-and-tell demonstrations. Of these six Ray consistently demonstrated both nonverbal and verbal PSI engagement factors through eye contact and responding thoughtfully to her audience. Further, a 10-week sampling of Twitter "good feeling" and "bad feeling" terms was collected and analyzed using Netlytic. Most of the terms found were associated with "Good Feelings." Analyzing the sub-categories of good and bad feelings words identified which terms were most often used to describe the show. Hypotheses were created to test for significant differences in the average number of terms. Implications of these findings for audiences of similar shows are discussed.

Keywords: parasocial interaction theory, direct address

The Interactive Medium of Television

As early as 1956 Horton and Wohl pointed out that broadcast media, especially television, had created "a new type of performer" (p. 216) who was particularly suited to induce and carry on a pseudo-relationship with her/his audience via the medium itself. In her study of how viewer identity is socially constructed through parasocial interactions (PSIs) with TV personalities Annese (2004) reinforced Horton and Wohl's findings and specified television's unique "communicative function" (p. 373) in relationship building. Giles (2002) wrote that an eclectic range of personalities have been the focus of audiences' attention, and, ultimately, PSIs, including comedians, shopping network MCs, soap opera figures, and many others. Recently, Madison et al. (2015) found that television has a particular "parasociability" (p. 259) that allows viewers to easily substitute the medium for actual in-person exchanges. The parasocial effect between viewers and media figures has been studied in a wide range of formats, from entertainment to news reporting to commercial, according to Liebers and Schramm (2019).

While the concept of parasocial relationships was defined by Horton and Wohl in 1956, modern media has changed significantly. Since that time, the world has gone from few homes having a television, to focusing on influencers through media such as podcasts, YouTubers, and platforms such as Facebook and Instagram. Madison, Wright, and Gaspard (2020) state "researchers can no longer ignore the functional power that parasocial relationships have in people's lives. It is critical that researchers recognize the power that parasocial relationships have in shaping public opinion among audiences" (p. 61). This focus on the criticality of research mentioned by Madison et al., is important for focusing on how PSRs affect both the influencer and the viewer in the role of business-building and sustaining a brand. While these are integral in building a brand, both positive or negative emotional and mental health effects of viewers ought to also be examined in these relationships. But how, to what extent, and what impact these relationships—even though they are one-sided, pseudo-relationships—remains to still be seen.

Television hosts are primed to develop parasocial relationships with their viewers due to their ability to engage audiences through visual and auditory modes. Understanding this engagement and how it results in connection with audiences is critical to audience-building and the ability of the influencer (in this case Rachael Ray) to build and sustain a profitable platform. Ray is ideal for this study due to her success in syndicated television, awards won, and the lifestyle brand she promotes through ancillary products such as cookware and her quarterly magazine. All of these avenues allow Ray to connect with viewers in a tangible way to that they may also mimic her portrayal of a quality life. Current research increasingly analyzes the broader topic of PSRs. While this is important, our study analyzes what is not often explored in PSRs; an individual influencer's brand and to what extent they connect with audiences interpersonally and nonverbally. Additionally, in what ways do they foster a sense of friendship, trust, and connectedness with the television audience? Analyzing these emotional items enhances existing literature through its specificity. Practical implications include supporting other influencers in developing PSRs through branding and audience engagement through fostering the aforementioned sense of friendship, trust, and connectedness with their audience.

By examining how Rachael Ray engages her audience through nonverbal and interpersonal communication, and how she fosters a sense of friendship, trust, and connectedness with them, this study exemplifies how Ray has developed highly successful PSRs, contributing to the literature by sharing specific techniques that influencers can use to develop PSRs and their media brand.

Literature Review

Para-social Interaction (PSI)

Horton and Wohl (1956) first introduced the concept of "para-social interaction" (p. 215). Their seminal work in creating this theoretical perspective adroitly explained all of the facets and inner workings of the pseudo-relationship that is possible in and across a variety of communication media, but which, according to the researchers, is most potent in the instantaneous media of radio and television. The para-social interaction is, of necessity, initiated by a media persona, who Horton and Wohl felt instigates and insinuates a transient relationship. However, the bulk of the ongoing illusory aspects of the relationship is subsequently reciprocated by audience members. Each TV viewer, for example, varies in her degree of *buy-in*, depending on how much each is invested in the overall illusion, that is, in the establishment and maintenance of the pseudo-interaction.

McQuail et al. (1972) validated Horton and Wohl's concept of para-social relationships in their study of TV quiz shows and radio and television programs. They specified that media use involves the "social interchange" (p. 144) of specific roles taken on by both sides – those in the programs and those in the audience. McQuail et al. found that the audience's role comes from gratification seeking, specifically, that of companionship. They described it as a "vicarious relationship with media personalities … as if he was on friendly terms with them, and as if they could stand in for real persons" (p. 157). Participants in their study indicated they experienced a "substitute companionship" (p. 157) with the TV characters, with some remarking "you feel you know them" (p. 157).

Alperstein (1991) also reinforced the concepts and work of Horton and Wohl in his study of pseudo-interpersonal relationships with celebrities appearing in TV ads. He asserted that television viewers can get a false sense that on-camera personas are speaking one-on-one with them and may even believe a celebrity is their friend or social acquaintance (see also Nisbett and Dunn, 2021). Further, according to Alperstein, the more an ad is seen over time, the stronger the illusory relationship may seem to grow (i.e., in the minds of some viewers), increasing the perception of connectedness, such as empathy with the celebrity's challenges or emulation of their success. Parasocial relationships may even occur with political ads where celebrities endorse a particular candidate (Nisbett & Dunn, 2021); however, engagement with the political message is "a key mediator between parasocial and outcome variables" (p. 34). Thus, Nisbett and Dunn emphasized the importance of the message itself.

Madison et al. (2015) closely connected parasocial interactions (PSIs) and parasocial relationships (PSRs) with a third concept, "imagined interactions (IIs)" (p. 259). Among the many aspects of IIs the researchers considered they "could" serve as a *substitution* [emphasis added] for real-life interaction" (p. 260). They aver the work of Caughey (1984, 1986) who found that audiences may create elaborate mental fantasies involving themselves and media personalities. Madison et al. suggested fantasies may compensate for the lack of reciprocation from a celebrity. Some fans may even try taking their imaginary interactions further and attempt to engage celebrities in person.

Madison et al. (2015) uncovered a number of specific parasocial behaviors, such as those described by Honeycutt (2010), including imagining dialogues with celebrities. Madison et al. point out that news audiences may even praise or criticize subjects of stories, even hoping to engage them in person. Other viewer mental activity that may enhance the parasocial effect include multiple re-remembering of situations seen on television, as well as trying to " . . . maintain relationships with both TV characters and storylines" (Madison et al., 2015, p. 266, referencing Madison & Porter, 2015). With regard to repetitive fantasies, Madison et al. references Caughey's (1984) findings that fans may concoct elaborate machinations that run the gamut of social interaction from casual to long-term imaginary relationships. In so doing they may construct "'arsenals' of cognitive scripts" (p. 266) for potential future use wherein they engage media personae in a variety of scenarios unlike those on screen, all the while knowing the unlikelihood of actually acting out the scripts. Madison et al. (2020) reports that while viewers may have these imagined interactions with television or online personalities they have never seen, these imagined interactions help "…people to assemble their thoughts, clarify viewpoisnt, and better understand their perspectives" (p. 53).

Some of the connection and engagement may result from interacting with reality television, for example, when the public is allowed to vote (Tian and Yoo, 2015). In their research on *The Biggest Loser* TV reality show, Tian and Yoo found that "the more frequently audience members view a show, the more likely they are to parasocially interact and identify with its cast members" (p. 5), which can impact their voting frequency and choices. This impact may even change feelings of hopelessness and future expectations in the lives of older adults, explain Chory-Assad and Yanen (2005) who found that "noting similarities with positive television performers (i.e., identification) may function as a means of escaping one's own unsatisfactory life" (p. 195). Similarly, Pan & Zeng (2019) found that "media viewers identify with media characters because both groups share certain salient characteristics" (p. 5). This identification as viewers may find connection in exploring interests like their PSR, which may increase well-being (Hoffner & Bond, 2022), Conversely, Madison et al. (2020) in their study of Alex Jones of *Infowars*, "found an association between imagining interpersonal conflict with Jones and diminished credibility of *Infowars*" (p. 61).

According to Kassing and Sanderson (2009), parasocial interaction involves interpersonal engagement on a variety of levels. This engagement includes activities such as sharing quotes (p. 194), expressing gratitude (p. 195), advice giving (p. 196), and play (p. 197). They concluded that these behaviors increase as "Internet communication technologies… afford users greater means for interacting" (p. 198). Sanderson and Truax (2014) found that people may become so caught up in their PSI/PSR with media personas that they place themselves into imagined personal narratives with the stars. This phenomenon has been further facilitated with the advent of social media and interactive social networks. “Parasocial interaction processes are more likely to occur when media viewers perceive media characters as more real and attractive, but can be simply switched or modified as media viewers reasonably expect or meet their similar characteristics in other media characters” (Pan & Zeng, 2019, p. 5).

Parasocial Interaction and Entertainment Television

Cohen (2004) found that early attitudes and beliefs about TV personalities are built into later perceptions of relationships with them and may influence behaviors related to perceived interactions and relationships. Schiappa et al. (2006) studied the effects of positively portrayed homosexual relationships in the media and how viewers’ reactions were enhanced toward those relationships by viewing the 8-season run of *Will and Grace*. They found that “ . . . parasocial interaction and viewing frequency are causally intertwined with each other in their relationship to levels of sexual prejudice toward gay men” (p. 10). As early as 1987 Rubin and Perse had signaled this causation in their research on soap opera viewing and audience responses to watching soaps. They concluded that soap operas offer “ . . . a relaxing way to escape more demanding interpersonal interaction . . . ” (p. 264) with those who did not view homosexual relationships as positive. This interaction and response to relationally-based media likely afforded viewers a place of entertainment and engagement without bias in a national culture that was on the cusp of viewing homosexual relationships in a more positive light, but a national culture which was not ready to accept them and their public displays of affection. Eyal and Cohen’s (2006) research on fans of *Friends* discovered that television personalities provide viewers with singular directed relationships and the intimacy they offer. They learned that viewers recognized favorite television characters as an extension of themselves, like friends but more familiar than acquaintances. The researchers noted that television can create a relationship that can be seen as a significant other, and media personalities are a substantial part of one's social network. In some examples viewers look to their media characters to serve as role models helping them develop goals in their lives modeling how they should see themselves and live their lives (Eyal & Cohen, 2006).

Conversely, in their study of disliked TV characters Dibble and Rosaen (2011) found that viewers' comprehension of a persona's action and situation can build emotions within a perceived relationship. It is common for viewers to experience sympathy, empathy, verbal, and nonverbal behavior during PSI engagement. Parasocial processing increases the relevance of the viewing experience for audience members and is necessary to achieve communicative-like reactions within the PSI process (Dibble & Rosaen, 2011). Comparing viewers' commitment to fictional television characters versus nonfictional television personalities Branch et al. (2013) found that "PSR partners are not dissimilar to real relationship partners" and that "it is reasonable to expect, then, that the experience of commitment to a PSR is not dissimilar to that of commitment to an interpersonal relationship (p. 20). Branch et al. conclude that PSRs "seem[s] to be virtually without bounds" (p. 20). Sanderson & Truax (2014) found that some people become so caught up in their PSI/PSR with media personas that they place themselves into personas narratives with these media stars. This has been made even easier with the advent of social media and interactive social networks. Madison et al. (2015) found similar connections between fans and characters of *Grey's Anatomy*, with some experiencing mourning at the death of a prominent character. They cite Dr. Debra Burton who stated, in part, "We invest in those people and they become a part of our lives" (p. 262).

Finally, Natale (2017) examined parasocial bereavement experienced by soap opera fans when their favorite shows were cancelled, comparing it to “the loss of a close friend” (p. 19). Media marketers seem to understand the durability and the longevity of parasocial relationships and may attempt to exploit the continued connection some viewers have with media persona who take on similar roles to those that were *lost*, wherein they became the viewer’s parasocial *friend*. Thus, fans who have lived vicariously through the persona’s life in a previous television program may subsequently follow them in their new roles and may continue their connection to the new character, as well as to the persona’s personal activities (Natale (2017).

The Unique Format of the Television Talk Show

From the research based on popular media, it is clear that affective cognition is present in fictional parasocial relationships, but even early research in the area of PSRs determined that they are present even in non-fictional media relationships such as news personae (Levy, 1979). In his study of TV news viewers Levy found that viewers engaged in parasocial interactions find “[reassurance in] a familiar, friendly ‘image’ of their intimates-at-a-distance” and that those who are active in news media relationships “experience a sense of order, belonging, and context from their relationship with news personae” (p. 78). Studies have focused on the particular characteristics of the television talk show that may readily draw viewers into an imaginary social relationship with the celebrity host. Annese (2004) studied the television talk show in depth and astutely explained that this format provides a very effective “space” (p. 371, 385) for perceived mediated interaction to occur for several concurrent factors. First, electronic media (being easily accessible and ever-present) tends to provide an interactive environment in which viewers can be inundated beyond capacity with ongoing, socially constructed messages relating themselves to others in the media world. Second, the medium of television readily transposes itself (i.e., crosses into) the reality of viewers' worlds and makes it seem that viewers are actually participating alongside and interacting with the talk show host. Third, viewers get a genuine sense of active "involvement" (p. 374, 382, 384) in the shows and can take on a number of virtual (i.e., imaginary) roles by constructing personalized interpretations.

Direct Address to the Audience.

Most pertinent for the current research are studies that focus specifically on direct address by the speaker to the media audience. Auter and Davis (1991) remind readers that this "breaking [of] the invisible 'fourth wall'" (p. 165) in television owes its creative origin to theater and also to early movies, mostly for comic relief and plot solutions in the latter. Their influential study of college students watching TV comedies with direct address to the audience found that the students' perceptions of the shows' "entertainment value and content sophistication" (p. 170) were significantly enhanced by direct address. This technique was well-liked and led to audience participation in the shows' content, even for older shows which were not known to them. Auter and Davis identified this as the "interest-involvement" (p. 170) process, in which an "interactive relationship redefines the normally passive relationship with a given show" (p. 170).

More recently, using *The Biggest Loser* reality TV show as a stimulus, Cummins and Cui (2014) distinguished between parasocial relationships (PSRs) and parasocial interactions (PSIs), the former being more slowly developed over time, and the latter regarding what takes place during the viewing of a show. They point out that research on the latter (PSIs), going back to Horton and Wohl (1956), successfully examined the formal features of programs using direct address, distinguishing between various styles of interactions that media personae (TV hosts) had with audiences. In particular, Cummins and Cui explained "bodily address" (pp. 726, 727) wherein the host looks at (via camera) and talks to the audience specifically, "verbal address" (pp. 726, 727), wherein video cutaways are inserted, but the host continues in voice-over mode, talking to the audience indirectly or even more directly using inclusive language, and "no address" (p. 727), wherein the host is obviously talking to someone else on the show. Of these three, the first two forms of address (bodily and verbal) have been found to facilitate a sense that the on-screen celebrity is aware in some sense of the viewer being there and is interacting by sharing thoughts and feelings, as in an interpersonal conversation. Bodily address, in particular, tends to foster a sense of emotional connection with the celebrity.

This emotional connectivity may be based on the degree to which hosts reveal personal real-life details Schlutz and Hedder (2022) "found that hosts that shared details about their own lives, were considered competent, authentic, and unpredictable were associated with greater parasocial relationships with listeners, which was then associated with listener behaviors and attitudes, thus, advertisers may also want to consider hosts that exhibit these qualities" (as cited in Fisher, 2024, p.13).

A Bright Spot in the TV Lineup: The Rachael Ray Show

Rachael Ray began her highly-popular, syndicated food and variety show on September 18, 2006, and garnered an average of 2.6 million views every day in its first season, making it very competitive with its daytime contemporaries. The show has been nominated for fifteen Emmy Awards and has won three Emmys (Rachael Ray, 2023). As of this writing *The Rachael Ray Show* is in its seventeenth season (Rosario, 2023), with over 2000 episodes (Rachael Ray, 2023) as a syndicated one-hour, semi-scripted cooking demonstration and pop-celebrity interview/talk show that also features short segments on self-help, public information, and letters from fans, among others. Each episode's 42 minutes of run time (Rachael Ray, 2023) features approximately 28% (about 12 minutes) of Rachael Ray preparing a full meal, broken down into three consecutive segments, the first two running approximately 5 minutes each, and the third, about 2 minutes. Two interview/talk segments make up about another 28% of each episode (about 5-7 minutes each). These may feature two separate guests or one more well-known guest, and may be in-person or online interviews. Finally, three or four miscellaneous, short segments typically finish out the show, comprising about 28% (about 3-4 minutes each) of the running time. The remaining show run time (about 13%, 6 minutes) consists of the intro, teases, next-show promos, and the wrap-up. Each of the primary show segments is further broken down into mini-segments on her YouTube channel (Rachael Ray Show) and her website (rachaelrayshow.com).

Throughout the cooking segments, in particular, and to a slightly lesser extent in the interview segments, Rachael makes frequent eye contact with her in-studio audience by looking directly at them and with at-home audiences through glances and gazes at the camera lens. Her nonverbal messaging is bright and positive, including smiling, making “small talk,” and self-deprecating humor. Her voice is soft and encouraging, and her approach is to invite both audiences into the discussion, whether shot in her New York TV studio or in her home kitchen.

Previous studies have explored PSRs on a macro-level of the topic with a generally broad focus, highlighting the prevalence of PSRs, but few PSR articles have been written about a specific personality and what characteristics help develop their PSR, making this an understudied aspect of PSRs, PSIs, and IIs. The unique focus of this study is Rachael Ray herself and her ability to connect with a wide variety of ages and interests, and namely, how she performs those things that develop connection.

The following research questions were investigated to further understand the nature of the interactive format used by Rachael Ray in the various segments of her show, and particularly, the cooking segments:

RQ1: Does Rachael Ray engage her television audience through nonverbal communication?

RQ2: Does Rachael Ray engage her television audience through interpersonal communication?

RQ3: In what ways does Rachael Ray foster a sense of friendship with her television audience?

RQ4: In what ways does Rachael Ray foster a sense of trust with her television audience?

RQ5: In what ways does Rachael Ray foster a sense of connectedness with her television audience?

Method

In order to test variations on null hypotheses 1 and 2 a 10-week sampling of Twitter terms was collected and analyzed using a Netlytic Tier-2 account. Two terms yielded the most utilitarian results and were further compared by running t-Tests and ANOVA tests: "Rachael Ray Show" (Condition A) and "Rachael Ray" (Condition B). The use of these two terms was associated with mostly "Good Feelings" (words) and a few "Bad Feelings" (words) within the texts of the tweets collected (see Results below). This study then looked deeper into the sub-categories categories of Good and Bad "feelings" words to identify which particular were most prevalent under each general term.

Hypotheses were created to test for significant differences in the average number of bad feeling terms across both time periods and conditions. These were operationalized using consecutive month-to-month periods, coded as continuous independent variables (IV). Hence, numerous search parameters (conditions) were set up using Netlytic for Twitter posts from December 2020 through the spring and late summer/early fall of 2021. A "Dictionaries" search under "Description" was run on each condition. After each analysis, many of the text terms under "Feelings: Bad" were recorded, and for each term, the per month average for the entire analysis run period was determined for each condition.

To test the hypotheses below various statistical analyses were run to determine the significance of the differences both within and between various conditions. A t-Test was used to compare means between two one-period search parameters. This was followed by ANOVA testing on each of the one-period parameters for differences between cells in the use of bad feeling terms. ANOVA also tested for significant differences within and between each of the multi-period conditions, regarding bad feeling terms and the time periods under each condition. The hypotheses and test results are reported thereafter.

Comparing the Conditions

In order to determine whether reaction to *The Rachael Ray Show* was more positive or negative overall, as well as whether particular good feeling or bad feeling terms were tweeted about the show, this study first sought to reject the following null hypotheses:

H_{01a} There will be no significant difference between the overall means of all bad feeling terms and the overall means of all good feeling terms used in conditions A and B.

H_{01b} There will be no significant difference between the means of specific bad feeling terms and the means of specific good feeling terms used *between* conditions A and B.

Next, this study sought to reject the following null hypotheses:

H_{02a} There will be no significant difference between the means of specific bad feeling terms used *within* condition A.

H_{02b} There will be no significant difference between the means of specific good feeling terms used *within* condition A.

H_{02c} There will be no significant difference between the means of specific bad feeling terms used *within* condition B.

H_{02d} There will be no significant difference between the means of specific good feeling terms used *within* condition B.

In seeking to answer RQ's 1 and 2, six distinct categories of parasocial engagement initiated by Ray were observed and counted in ten randomly sampled episodes from seasons 15 and 16 (n = 20) of *The Rachael Ray Show*, for a total of 20 episodes. Briefly, these were articulated as 1) making eye contact, 2) responding thoughtfully, 3) verbally building rapport, 4) making herself appreciable, 5) follow-along descriptions, and 6) show-and-tell demonstrations. Moreover, in order to test variations on null hypotheses 3, the 20 randomly selected episodes were examined for attempts made by the host, Rachael Ray, to intentionally elicit parasocial interaction with the TV and/or studio audience. Comparisons of averages between total observations in each of these six categories for the sampled episodes were analyzed for statistical significance by running t-Tests and ANOVA tests, in an attempt to reject the null hypotheses, as stated below.

Specifically, this study sought to reject the following null hypotheses:

H_{03a} There will be no significant difference between the mean number of occurrences in any of the six categories within the two seasons studied overall.

H_{03b} There will be no significant difference between the mean number of occurrences in any of the six categories within the individual episodes of the two seasons studied.

Results

The two terms "Rachael Ray Show" and "Rachael Ray" were associated with mostly "good" feelings and a few "bad" feelings within the texts of the tweets collected. Looking deeper into the sub-categories of good "feelings" and bad "feelings" words for each of these general terms yielded several interesting sub-categories of words that were used, although the good feeling terms clearly outnumbered the bad terms, which further reinforces the concept that PSI works most effectively as a positive bond between personality and audience member. Unfortunately, H_{01a} could not be rejected at this point in the research, in comparing the overall means of the use of “good feeling” and “bad” feeling terms using a number of statistical tests, including two-sample and paired two-sample t-Tests on the means, as well as single-factor ANOVA and two-factor ANOVA without replication. Each of these tests yielded p-values too high to reject H_{01a}.

H_{01b} was more specific, in that it assumed no significant differences among *specific* bad feeling and good feeling terms tweeted in response to the different conditions, which are slightly differentiated Twitter search terms, specifically, Rachael Ray and *The Rachael Ray Show*. The hypothesis was strongly rejected using two ANOVA *between*-condition (columns) comparisons of the means of various bad feeling and good feeling terminology used between the conditions, specifically, the ANOVA single-factor analysis of *between*-condition comparison ($p < .05$) (see Table 1) and the ANOVA two-factor analysis (without replication) of *between*-condition (columns) comparison ($p < .01$). The latter ANOVA analysis also indicated significant differences *within*-conditions, comparing specific bad and good feelings ($p < .06$) (see Table 2).

Further, comparing seven specific good feeling terms given in response to the two conditions using ANOVA single-factor analysis indicated significant *between*-condition differences ($p < .02$) (see Table 3). Expanding the analysis to include ten specific good feeling terms given in the two conditions, ANOVA two-factor analysis (without replication) indicated even more significant *between*-condition differences ($p < .01$). This analysis also indicated significant differences *within*-condition, comparing the ten good feeling terms to each other ($p < .04$) (see Table 4).

Hypotheses H_{02a} and H_{02b} could not be accepted or rejected due to insufficient data (responses) to the condition search term "*Rachael Ray Show*." Nevertheless, both H_{02c} and H_{02d} were strongly rejected using comparisons of the means of specific bad and good feeling terms tweeted in response to the condition search term "Rachael Ray." Specifically, one-tailed and two-tailed t-Tests had p-values of .01 and .02, respectively, indicating significant differences among both good feeling and bad feeling words (see Table 5).

Both H_{03a} and H_{03b} were strongly rejected using both ANOVA single-factor analysis and ANOVA two-factor analysis (without replication) of *between*-categories comparisons of means, ($p < .001$), ($p < .001$), respectively (see Tables 6 and 7). This finding indicates significant differences between the categories of PSI using both single-factor analysis and two-factor analysis, but there was no significant difference between the rows (episodes) (see Table 7). Furthermore, the one-tailed t-Test had a p-value of .05 indicating a significant difference between means in the first two categories of parasocial engagement by Rachael Ray, namely: listening while making eye contact and thoughtfully responding to questions asked (see Table 8). Thus, for RQ's 3-5 it is found that, at least with these two types of parasocial interaction, Rachael Ray does indeed make an effort to foster a sense of friendship, trust, and connectedness with her live (in studio) and television audience.

Discussion

This study utilized Parasocial Interaction theory (PSI) to examine the presence of relationship-building techniques between talk show host and cooking celebrity Rachael Ray and her television audience. PSI theory explains that media personalities, particularly those who are the primary focus of TV show, such as the main host or emcee, are the figures with whom viewing audiences are most likely to identify with, feel a connection to, and even assume some degree of affiliation with and responsiveness from the media figure. PSI theory also indicates that hosts/MC's emcees are primarily responsible for creating the feeling of an imaginary relationship with viewers because they actively seek some level of participation from their anonymous followers.

Analysis of randomly selected episodes from two recent seasons of Rachael Ray's eponymous, syndicated interview/demonstration show have revealed that Ray purposively engages her TV audience (as well as the studio audience) with a variety of verbal and nonverbal techniques. Of primary importance, Ray directly addresses the at-home (and studio) audience directly, initiating a familial connection, while simultaneously making eye contact (via the camera lens) in a way that closely resembles intimate conversation between close acquaintances. It is through the combination of direct address and eye contact that Ray's gregarious nature is initiated, while her tone, down-to-earth style, and obsequious posturing indicate complete openness and invite a give-and-take interaction from viewers. Ray augments her direct address with questions for her unseen, non-present respondents and inside jokes and self-deprecating humor that enhance the feeling that the on-screen presentation has been tailored personally to each of them, rather than aggregately.

Additionally, it is likely that Ray's intentionally interactive approach has been an important factor in the show's overwhelming success. Twitter responses to *The Rachael Ray Show* have indicated overwhelmingly good feelings associated with the show, as opposed to bad feelings. These feelings undoubtedly lead to increased fan-ship with the media figure and an imagined bonding with Ray. Bore (2009) employed focus group analysis in a study of the British version of the hit television sitcom "The Office." Bore found that the participants acknowledged the perception of realism in the show, regardless of their enjoyment or displeasure with the humor and "mockumentary[sic]" (p. 36) format. Further, Kocela (2009) found that the pseudo-interview elements when characters in "The Office" address the camera "construct an emotional truth about their subjects that is predicated on the strategic positioning of the camera as a silent, impartial recorder of events" (p. 166).

Readers and viewers may ask why parasocial relationships should matter or be considered. In the case of the Rachael Ray empire, one might liken her personality to an Oprah-esque personality who had a tremendous impact on primary female viewers through her daily talk show, book clubs, and even her own television station which was aptly named *OWN*. While Ray has not attained the same level of popularity that Oprah enjoyed, the case could easily be made that she impacts her (again primarily female) audience in a positive way. "Social media influences invite their followers into their day-to-day lives, so followers feel they are pseudo-friends with the influencers" (Perez, Schultz and Landreville, 2024, p. 3).

The age of typical adulthood behaviors is growing later each decade (Hill & Redding, 2021: Hayford & Furstenberg, 2008: Twenge & Park, 2019). Due to the delay of these behaviors, young adults will undoubtedly look for those to teach them the activities of daily living they may have missed in this extension. Parasocial relationships such as those with the warm and inviting Rachael Ray can have a positive affect on adults of all ages and possibly lessen some of the anxiety of living in an adult world for those who are quite unsure of what this looks like. In a world where adults of all ages are seeking the previously referred to qualities that she exhibits of a sense of friendship, trust, and connectedness, what more could be asked of a parasocial relationship?

Limitations and future directions

Since this research focused solely on the *Rachel Ray Show*, the focus was limited to one media personality. The research does not show whether or not another media personality who displays the same characteristics as Ray would have the same effect on her audience as she does. In summary, the question regarding this is, *"Is it Ray's intentional behavior or her intrinsic self that draws her audience into the parasocial relationship?"*

Future research could the research questions in this article to see how they parlay to the PSR experience that other media personalities have with their audiences. Additionally, a research focus could be conducted that considers the positive or negative affect of integrating parasocial relationships with persons who exhibit anxious attachment styles in therapeutic/behavior modification settings. Citing Pavelko and Myirck (2020), Fisher (20224) reinforces while many use podcasts for enjoyment. Fisher refers to listeners of the "My Favorite Murder" podcast who "…perceived social support" (p.3) by hosts who "encouraged openness about mental health" (p. 3). While this paper focuses on Rachael Ray and the medium of television, further research could be conducted on audio forms on entertainment regarding PSRs and mental health. Additionally, behavior modification could be studied to determine whether embodying the characteristics of the beloved personality makes a difference in confidence and/or lessens an anxious attachment style.

References

Alperstein, N. M. (1991). Imaginary social relationships with celebrities appearing in television commercials. *Journal of Broadcasting & Electronic Media*, *35*(1), 43. https://doi-org.glbvvproxy.enmu.edu/10.1080/08838159109364101

Annese, S. (2004). Mediated identity in the parasocial interaction of TV. *Identity*, *4*, 371–388.

Auter, P. J., & Davis, D. M. (1991). When characters speak directly to viewers: Breaking the fourth wall in television. *Journalism Quarterly, 68,* 165–171.

Bore, I.-L. K. (2009). Negotiating generic hybridity: Audience engagement with The Office. Continuum: Journal of Media & Cultural Studies, 23(1), 33–42. https://doi-org.glbvvproxy.enmu.edu/10.1080/10304310802570882

Branch, S. E., Wilson, K. M., & Agnew, C. R. (2013). Committed to Oprah, Homer, or House: Using investment model to understand parasocial relationships. *Psychology of Popular Media Culture*, *2*, 96–109.

Caughey, J. L. (1984). Imaginary social worlds: A cultural approach. Lincoln, NE: University of Nebraska Press.

Caughey, J. L. (1986). Social relations with media figures. In G. Gumpert & R. Cathcart (Eds.), Inter/media: Interpersonal communication in a media world (3rd ed., pp. 219–252). New York, NY: Oxford University Press.

Chory-Assad, R. M., & Yanen, A. (2005). Hopelessness and loneliness as predictors of older adults' involvement with favorite television performers. *Journal of Broadcasting & Electronic Media*, *49*(2), 182–201. https://doi org.glbvvproxy.enmu.edu/10.1207/s15506878jobem4902_3

Cohen, J. (2004). Parasocial breakup from favorite television characters: The role of attachment styles and relationship intensity. *Journal of Social and Personal Relationships*, *21*, 187–202.

Cummins, R. G., & Cui, B. (2014). Reconceptualizing address in television programming: The effect of address and affective empathy on viewer experience of parasocial interaction. *Journal of Communication*, *64*(4), 723–742. https://doi-org.glbvvproxy.enmu.edu/10.1111/jcom.12076

Dibble, J. L., & Rosaen, S. F. (2011). Parasocial interaction as more than friendship: Evidence for parasocial interactions with disliked media figures. *Journal of Media Psychology, 23*(3), 122-132.

Eyal, K., & Cohen, J. (2006). When Good "Friends" Say Goodbye: A Parasocial Breakup Study. *Journal of Broadcasting & Electronic Media*, *50*(3), 502–523. https://doi.org/10.1207/s15506878jobem5003_9

Fisher, N. (2024). But First, A Word from Our Sponsor: Engagement with Podcasts and their Advertising. Southwestern Mass Communication Journal, *40*(1), 1-16. doi: https://doi.org/10.58997/zjmenf09

Giles, D. C. (2002). Parasocial Interaction: A Review of the Literature and a Model for Future Research. *Media Psychology*, *4*(3), 279–305. https://doi.org/10.1207/S1532785XMEP0403_04

Hayford SR, Furstenberg FF Jr. Delayed Adulthood, Delayed Desistance? Trends in the Age Distribution of Problem Behaviors. J Res Adolesc. 2008 Jun 1;18(2):285-304. doi: 10.1111/j.1532-7795.2008.00561. PMID: 19633730; PMCID: PMC2714655.

Hill, N., & Redding, A. (2021). The End of Adolescence: The Lost Art of Delaying Adulthood. Harvard University Press.

Hoffner, C.A. & Bond, B.J. (2022). Parasocial relationships, social media, & well-being, *Current Opinion in Psychology*, *45*, ISSN 2352-250X, https://doi.org/10.1016/j.copsyc.2022.101306.

Horton, D., & Wohl, R. R. (1956). Mass Communication and Para-Social Interaction: Observations on Intimacy at a Distance. *Psychiatry, 19*(3), 215-229.

Honeycutt, J.M. (Ed.). (2010). Imagine that: Studies in Imagined Interaction (pp. 1–14). Cresskill, NJ: Hampton Press, Inc.

Kassing, J., & Sanderson, J. (2009). "You're the Kind of Guy That We All Want for a Drinking Buddy": Expressions of Parasocial Interaction on Floydlandis.com. *Western Journal of Communication*, *73*(2), 182–203. https://doi-org.glbvvproxy.enmu.edu/10.1080/10570310902856063

Kocela, C. (2009). Cynics Encouraged to Apply: The Office as Reality Viewer Training. *Journal of Popular Film & Television*, *37*(4), 161–168. https://doi-org.glbvvproxy.enmu.edu/10.1080/01956050903227886

Levy, M. R. (1979). Watching TV News as Para-Social Interaction. *Journal of Broadcasting, 23,* 69–80.

Liebers, N., & Schramm, H. (2019). Parasocial Interactions and Relationships with Media Characters-An Inventory of 60 Years of Research. *Communication Research Trends*, *38*(2), 4–31.

Madison, T. P., & Porter, L. V. (2015). "The People We Meet." *Imagination, Cognition and Personality: Consciousness in Theory, Research, and Clinical Practice*, *35*(1), 47–71. https://doi.org/10.1177/0276236615574490.

Madison, T. P., Porter, L. V., & Greule, A. (2015). Parasocial Compensation Hypothesis:

Predictors of Using Parasocial Relationships to Compensate for Real-life Interaction. *Imagination, Cognition, and Personality: Consciousness in Theory, Research, and Clinical Practice*, *35*(3), 258–279.

Madison, T.P., Wright, K., Gaspard, T. (2020). My Superpower is Being Honest: Perceived Credibility and Parasocial Relationships with Alex Jones. *Southwestern Mass Communication Journal, 36*(1), 50-64. doi: https://doi.org/10.58997/smc.v36i1.79

McQuail, D., Blumer, J. G., & Brown, J. R. (1972). The Television Audience: A Revised Perspective. In D. McQuail (Ed.) *Sociology of mass communications: Selected readings* (pp. 135–165). Harmondsworth: Penguin.

Natale, D. W. (2017). Parasocial Bereavement: Where Do They Turn When their Soap Opera Friends Depart? *Florida Communication Journal*, *45*(1), 19–33.

Nisbett, G., & Schartel Dunn, S. (2021). Reputation Matters: Parasocial Attachment, Narrative Engagement, and the 2018 Taylor Swift Political Endorsement. *Atlantic Journal of Communication*, *29*(1), 26–38. https://doi-org.glbvvproxy.enmu.edu/10.1080/15456870.2019.1704758

Pan, P. and Zeng, L (2019). The Impact of Ethnicity Match on Athlete Endorsement: A Parasocial Idenitifcation Persepctive. *Southwestern Mass Communication Journal,39* (1). 1-15. doi: https://doi.org/10.58997/smc.v39i1.114

Perez, A., Schultz, C., & Landreville, K. (2024). The Try Guys Try Cheaing: Social Media Influencers in Crisis. *Southwestern Mass Communication Journal,40(1), 1-19.* https://doi.org/10.58997/be6hzy42

Rachael Ray (talk show). (2023, May 1). In *Wikipedia*.

https://en.wikipedia.org/w/index.php?title=Rachael_Ray_(talk_show)&oldid=115247494
9

Rosario, A. D. (2023, March 7). Rachael Ray is ready 'to move on' from daytime TV and already cooking up new ventures. *Los Angeles Times*. https://www.latimes.com/entertainment-arts/tv/story/2023-03-07/rachael-ray-ending-talk-show-daytime-tv-free-food-studios

Rubin, A. M., & Perse, E. M. (1987). Audience Activity and Soap Opera Involvement: A Uses and Effects Investigation. *Human Communication Research, 14,* 246–268.

Sanderson, J., & Truax, C. (2014). "I hate you Man!": Exploring Maladaptive Parasocial Interaction Expressions To College Athletes Via Twitter. *Journal of Issues in Intercollegiate Athletics*, *7*, 333-351.

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2006). Can One TV Show Make A Difference? Will & Grace and the Parasocial Contact Hypothesis. *Journal of Homosexuality*, *51*, 15–37.

Tian, Y., & Yoo, J. H. (2015). Connecting With The Biggest Loser : An Extended Model of Parasocial Interaction and Identification in Health-Related Reality TV Shows. *Health Communication*, *30*(1), 1–7. https://doi.org/10.1080/10410236.2013.836733 org.glbvvproxy.enmu.edu/10.1080/10410236.2013.836733

Twenge, J. M., & Park, H. (2019). The Decline in Adult Activities Among U.S. Adolescents, 1976–2016. Child Development, 90(2), 638-654. https://doi.org/10.1111/cdev.12930

Table 1. Single-Factor ANOVA comparison of Between-Condition Means of Various Good and Bad Feelings

Source/Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	68.05556	1	68.05556	6.524634	0.021216	4.493998
Within Groups	166.8889	16	10.43056			
Total	234.9444	17				

Table 2. Two-Factor ANOVA comparison of Between and Within Condition Means of Specific Good and Bad Feelings

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	127.4444	8	15.93056	3.230986	0.058644	3.438101
Columns	68.05556	1	68.05556	13.80282	0.005912	5.317655
Error	39.44444	8	4.930556			
Total	234.9444	17				

Table 3. Single-Factor ANOVA comparison of Between-Condition Means of Specific Good Feelings

Source/Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	82.57143	1	82.57143	7.846154	0.016012	4.747225
Within Groups	126.2857	12	10.52381			
Total	208.8571	13				

Table 4. Two-Factor ANOVA comparison of Between and Within Condition Means of Expanded List of Good Feelings

Source/Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	142.2727	10	14.22727	3.280922	0.037245	2.978237
Columns	69.13636	1	69.13636	15.9434	0.002547	4.964603
Error	43.36364	10	4.336364			
Total	254.7727	21				

Table 5. t-Test: Two Sample for Means of Specific Good and Bad Feelings for Condition "Rachael Ray"

	Variable 1	Variable 2
Mean	1.5	8.5
Variance	0.333333	11
Observations	4	4
df	3	
t Stat	-4.15862	
P(T<=t) one-tail	0.012643	
t Critical one-tail	2.353363	
P(T<=t) two-tail	0.025286	
t Critical two-tail	3.182446	

Table 6. Single-Factor ANOVA comparison of Between-Categories Means of PSI Engagement

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	52.94167	5	10.58833	7.140314	7.61E-06	2.293911
Within Groups	169.05	114	1.482895			
Total	221.9917	119				

Table 7. Two-Factor ANOVA comparison of Between-Categories and Rows (Episodes) Means for PSI Engagement

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	21.15833	19	1.113596	0.715332	0.794793	1.69707
Columns	52.94167	5	10.58833	6.801544	1.84E-05	2.310225
Error	147.8917	95	1.556754			
Total	221.9917	119				

Table 8. t-Test: Two-sample for Overall Means Between Categories 1 and 2 of PSI Engagement

	Variable 1	Variable 2
Mean	4.2	3.6
Variance	1.536842	0.989474
Observations	20	20
df	36	
t Stat	1.688194	
P(T<=t) one-tail	0.05001	
t Critical one-tail	1.688298	