ABSTRACT

ISABELLA GYIMAH, Effectiveness of Interventions for Girls in STEM: A Meta-Analytic Review within a Stereotype Threat Framework. (Under the direction of Dr. Isaac L. Woods).

Girls remain underrepresented in science, technology, engineering, and mathematics (STEM) fields, in part due to the persistent influence of stereotype threat; the fear of confirming negative societal beliefs about one's group. This meta-analysis examined the effectiveness of interventions designed to improve STEM outcomes among girls. Drawing on the Social Identity Framework (Tajfel & Turner, 1979), the study reviewed 12 peer-reviewed articles involving controlled experimental or quasi-experimental designs. Outcomes analyzed included STEM knowledge, interest, identity, self-esteem, and perception change. Hedges' *g* was used to compute effect sizes, accounting for small sample sizes, and heterogeneity was assessed using the Q statistic. Results showed small to moderate positive effects across most outcomes. For STEM knowledge, the pooled effect size was $g \approx 0.43$, indicating a modest improvement post-intervention. STEM interest and self-esteem outcomes also demonstrated favorable gains ($g \approx 0.38$ and $g \approx 0.18$ respectively), while STEM identity showed a slightly smaller effect ($g \approx 0.10$). Only one study reported quantifiable data on perception change, with a large effect ($g = -0.92$), suggesting significant stereotype reduction (Zhao et al., 2018). Heterogeneity was significant, highlighting some variability in study designs, contexts, and intervention types. Overall, the findings support the effectiveness of interventions in enhancing girls' STEM engagement and mitigating negative psychological impacts and highlights the need for early intervention. These insights are critical for designing evidence-based, equitable STEM education strategies that foster long-term inclusion.

Effectiveness of Interventions for Girls in STEM: A Meta-Analytic Review within a Stereotype Threat Framework

by
Isabella Gyimah

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Arts

Psychology

Raleigh, North Carolina
2025

APPROVED BY:

____
Isaac L. Woods, Ph. D

Committee Chair

____
Kelly Lynn Mulvey, Ph. D

____
Jennifer Burnette, Ph. D

BIOGRAPHY

Isabella Gyimah earned her bachelor's degree in psychology with Linguistics from the University of Ghana, graduating among the top ten students in her cohort. She is currently a third-year doctoral student in the School Psychology program at the North Carolina State University (NCSU). As part of her training, she completed school-based practicum placements at Knightdale Elementary School and Holly Springs Elementary School, where she supported academic and behavioral interventions for diverse student populations. Isabella is currently completing an advanced practicum at the Raleigh TEACCH Center, where she contributes to autism assessments and provides evidence-based interventions to children and families. Her research interests center on the intersection of educational equity, psychological science, and data-driven practices. She is passionate about conducting cutting-edge research that informs inclusive educational practices and addresses the needs of culturally and linguistically diverse children.

TABLE OF CONTENTS

CHAPTER 1 .... *1*

INTRODUCTION *1*

Literature Review 7

Theoretical Framework 10

Moderating Variables 13

The Current Study 16

Outcome Variables 17

CHAPTER 2 .... *22*

METHODOLOGY *22*

Search Strategy for Identifying Relevant Studies 22

Data Analysis 24

CHAPTER 3 .... *28*

RESULTS *28*

What is the Efficacy of Interventions in Overcoming Stereotype Threat and Improving STEM Outcomes? 28

Does the Effectiveness of STEM Interventions for Girls Vary Based on Grade Level, Intervention Type, Delivery Agent, or Origin of Study 32

Moderator Analysis 32

CHAPTER 4 .... *35*

DISCUSSION *35*

Cultural and Racial Analysis 42

Interpreting Intervention Effects Through the Lens of Social Identity Theory 44

Measures Used and Psychometric Considerations 45

Recommendations for Practice: Enhancing Girls' Engagement and Achievement in STEM 46

Limitations and Future Directions 50

Conclusion 52

References *53*

List of Tables *69*

Table 1.1 Description of Included Studies. 69

Table 1.2 Outcome Measures and Effect Sizes of Included Studies 71

Table 1.3 Distribution of Studies by Moderator Variables 73

Table 1.4 Reported Outcome Variables Across Studies 74

Table 1.5 Studies Excluded from the Meta-Analysis After Full Text Review and Reasons for Exclusion 75

Appendix *77*

Appendix A 77

LIST OF TABLES

TABLE 1.1 Description of Included Studies……………………………………………...….69

TABLE 1.2 Outcome Measures and Effect Sizes of Included Studies………………………….71

TABLE 1.3 Distribution of Studies by Moderator Variables…………………...…………....73

TABLE 1.4 Reported Outcome Variables Across Studies…………………...……………....74

TABLE 1.5 Studies Excluded from the Meta-Analysis After Full Text Review and Reasons for Exclusion……………………………………………….………………………………...75

CHAPTER 1

INTRODUCTION

Stereotype threat emerged as a concept from social and cognitive psychology, growing out of research on the psychological effects of negative societal stereotypes. Steele and Aronson (1995) first introduced the term in their seminal study, which demonstrated that Black students performed worse on standardized tests when their racial identity was made salient, suggesting that the fear of confirming negative stereotypes could impair performance. This foundational research built upon earlier theories of self-fulfilling prophecy (Rosenthal & Jacobson, 1968) and attributional ambiguity (Crocker & Major, 1989), which explored how individuals internalize societal biases. Subsequent studies expanded the concept to gender and STEM fields, notably Spencer, Steele, and Quinn (1999), who found that women underperformed on math tests when reminded of gender stereotypes. Over time, stereotype threat research has highlighted its broader implications, including reduced academic persistence (Beasley & Fischer, 2012), decreased interest in stereotyped domains (Good, Aronson, & Harder, 2008), and long-term impacts on career choices (Ceci, Williams, & Barnett, 2009). Today, stereotype threat is understood as a key psychological barrier that reinforces systemic inequities, shaping both individual behaviors and institutional outcomes (Schmader & Johns, 2003).

Steele and Aronson (1995), describe stereotype threat as the psychological pressure individuals experience when they fear confirming negative beliefs about their social group. This fear can negatively affect performance in different situations. This phenomenon is also relevant in academic settings, where girls often contend with pervasive stereotypes suggesting inferior aptitude in math and science (Spencer, Steele, & Quinn, 1999). These stereotypes not only undermine self-confidence but also contribute to the persistent gender gap in STEM (science, technology, engineering, and mathematics) fields (National Science Foundation [NSF], 2021). Historically, biases portraying women as less capable in STEM have been reinforced through media, educational practices, and societal messaging (Eccles, 1987). Research shows that stereotype threat can affect performance by diverting cognitive resources necessary for task completion, such as working memory and attention (Schmader & Johns, 2003). For instance, when girls are told “girls are not as good in math as boys” before taking a math test, their performance tends to decline significantly compared to conditions where such stereotypes are not primed (Good, Aronson, & Harder, 2008). These findings emphasize the pervasive nature of stereotype threat and its capacity to inhibit academic potential in areas where stereotypes are most entrenched. The implications of stereotype threat for girls in math and science are far-reaching. In classrooms, subtle cues, such as most math and science teachers being males or biased language in textbooks, can inadvertently trigger stereotype threat (Cheryan et al., 2009). This not only affects test performance but also shapes girls’ long-term attitudes toward STEM disciplines. For example, girls exposed to environments where stereotypes about their math abilities are reinforced are less likely to pursue advanced coursework or careers in these fields (Ceci, Williams, & Barnett, 2009). Consequently, addressing stereotype threat is critical for fostering gender equity in STEM education and beyond.

Educational contexts play a pivotal role in mitigating or exacerbating the effects of stereotype threat. Teachers' expectations and interactions with students can either reinforce or challenge gender stereotypes (Tiedemann, 2002). For instance, when teachers attribute boys' success in math to natural ability but girls' success to effort, they inadvertently reinforce the stereotype that math competence is innate for boys but not for girls (Eccles, 2011). Such biases, often unconscious, can influence students' self-perceptions and aspirations, particularly in high-stakes academic settings. In addition to its direct effects on performance, stereotype threat contributes to broader disparities in STEM participation. According to the American Association of University Women (AAUW, 2020), women constitute only 28% of the STEM workforce, with even lower representation in fields like engineering and computer science. This underrepresentation is not solely a result of individual choices but is deeply rooted in systemic barriers, including stereotype-driven biases that discourage girls from pursuing STEM pathways. Early exposure to stereotype threat can deter girls from identifying with STEM disciplines, resulting in a self-reinforcing cycle of underrepresentation and stereotype perpetuation (Correll, 2001). From an early age, girls receive implicit and explicit messages that STEM fields are better suited for boys, shaping their academic interests and self-perceptions (Eccles, 2009). These messages come from various sources, including teachers, parents, media, and cultural norms, which influence girls' confidence in their abilities and willingness to pursue STEM-related challenges (Nosek et al., 2002). Research suggests that when young girls perceive STEM as a domain in which they are less likely to succeed, they may disengage from these subjects, leading to fewer advanced coursework enrollments and decreased aspirations for STEM careers (Ceci & Williams, 2010). Moreover, a lack of female role models in STEM further reinforces the notion that these fields are unwelcoming or unsuitable for women, exacerbating the cycle of underrepresentation (Cheryan et al., 2017). This systemic issue contributes to the persistence of gender disparities in STEM fields, highlighting the urgent need for interventions that challenge stereotypes and foster a sense of belonging among girls in these disciplines. To address stereotype threat, it is essential to identify interventions that effectively improve the performance of girls in math and science. Unfortunately, limited empirical research has systematically evaluated the relative effectiveness of specific strategies designed to mitigate stereotype threat (Spencer et al., 2016). While existing studies have explored approaches such as role model exposure, and creating stereotype-safe environments, much of the focus has been on isolated interventions, with little attention given to how moderating variables, such as age, cultural context, type of intervention, delivery agent; influence the effectiveness of these interventions (Walton & Cohen, 2007; Steele, 2010). Thus, the purpose of this study will be to identify the most effective STEM intervention for combating stereotype threat, analyze how moderating variables, such as age, cultural context, type of intervention, delivery agent, influence the effectiveness of these interventions and identify areas for future research to inform evidence-based practices.

Background

Interventions to combat stereotype threat and improve STEM outcomes among girls have yielded encouraging outcomes, yet their effectiveness across diverse contexts remains inadequately explored. Research shows that interventions that focus on effort rather than innate ability have been linked to improved math performance and increased interest in STEM fields among girls (Good et al., 2003). Additionally, interventions that expose girls to positive role models in STEM have demonstrated significant potential. Research by Betz and Sekaquaptewa (2012) highlights that such role models not only challenge stereotypes but also inspire girls to envision themselves as capable and successful in STEM fields. Another impactful evidence-based strategy involves creating stereotype-safe environments. This includes anonymized testing methods, which reduce the salience of social identity, and curricular reforms that emphasize the contributions of women and other underrepresented groups in STEM (Beasley & Fischer, 2012). Research also underscores the importance of fostering a sense of belonging in academic settings.

Walton and Cohen (2007) found that students who feel accepted and valued within their academic environments are less likely to internalize negative stereotypes, thereby improving their performance and persistence in STEM. Additional interventions, such as female-only STEM workshops and classrooms, have been introduced to create supportive environments where girls can explore math and science free from gender-based pressures. Studies suggest that these gender-segregated settings can provide a space for girls to build confidence, enhance collaboration, and develop a stronger STEM identity (Riegle-Crumb et al., 2011). Professional development workshops for educators have also been implemented to train teachers on reducing implicit biases, using inclusive language, and promoting equitable classroom practices (Banaji & Greenwald, 2013). Furthermore, programs that incorporate peer mentoring and collaborative problem-solving activities have been shown to counteract stereotype threat by fostering community and shared success among students (Dasgupta, 2011).

Despite the promise of these interventions, a significant gap remains in understanding the contextual factors that influence their effectiveness across educational settings. Much of the existing research evaluates interventions in isolation, making it difficult to identify which approaches are most impactful for improving outcomes among girls in STEM. For instance, while role model-based interventions and curriculum modifications have been widely studied, few comparative studies examine how the type of intervention; whether it is values-affirmation, role model exposure, or growth mindset training, affects different outcomes. Similarly, the delivery agent, such as whether interventions are facilitated by teachers, researchers, or peers, may shape how messages are received and internalized, yet this remains underexplored (Walton & Cohen, 2011). The grade level at which an intervention is implemented is another critical consideration, as stereotype threat may emerge or intensify during key developmental transitions such as middle school (Eccles et al., 1993). Additionally, the study origin, including whether research is conducted in the United States or other cultural contexts, may influence generalizability and applicability of findings due to differences in gender norms and educational systems (Cheryan et al., 2015). These moderating variables are essential for understanding not only whether interventions work, but also for whom, under what conditions, and through which mechanisms they exert their effects.

This study addresses existing gaps by conducting a meta-analytic review of interventions aimed at mitigating stereotype threat and improving STEM outcomes among girls. Specifically, it examines the overall effectiveness of these interventions relative to each other while exploring key moderating variables such as intervention type, delivery agent, grade level, and study origin that may influence their impact. By analyzing how these contextual factors shape intervention outcomes, the study identifies the most effective strategies for reducing stereotype threat and promoting gender equity in STEM. The findings offer valuable insights for educators, policymakers, and researchers working to develop inclusive and evidence-based practices that support girls' success in STEM disciplines (Cheryan et al., 2017; UNESCO, 2021). Addressing stereotype threat is not merely an issue of academic achievement but a critical step toward broader social equity. As global demand for STEM talent continues to rise, fostering gender diversity in these fields is essential for driving innovation, ensuring economic competitiveness, and promoting social progress (Cheryan et al., 2017; UNESCO, 2021). By focusing on the psychological and contextual barriers that limit participation, this research contributes to a growing body of work aimed at building equitable and empowering educational environments for all learners.

Literature Review

A substantial body of research demonstrates that stereotype threat undermines girls' performance in math and science, perpetuating gender disparities in these fields. However, while some interventions have shown promise in mitigating stereotype threat, questions remain regarding the most effective strategies and their application across diverse contexts. This literature review synthesizes the current research on stereotype threat and interventions aimed at reducing its impact on girls in STEM, while also identifying gaps that this study seeks to address.

Historical Barriers and Systemic Inequities in Women's Representation in STEM

The underrepresentation of women in STEM fields is rooted in historical systemic barriers that have limited access to education, professional opportunities, and societal recognition of women (Eccles, 1987). In the early 20th century, women faced explicit exclusion from higher education and scientific careers, with institutions often barring them from enrolling in advanced coursework or attaining faculty positions (Rossiter, 1982). Even when women gained access, they encountered significant biases that reinforced traditional gender roles, discouraging them from pursuing STEM careers (Fox, 2001). The mid-20th century saw structural policies, such as restrictive hiring practices and unequal pay, that further marginalized women in science and engineering (Long & Fox, 1995). Moreover, cultural stereotypes associating scientific and mathematical ability with men often reinforced through educational curricula and media shaped early socialization processes, dissuading girls from developing strong STEM identities (Eccles, 1987). These barriers were perpetuated by implicit biases in academic institutions, where male-dominated networks and mentorship structures limited women's professional advancement (Etzkowitz, Kemelgor, & Uzzi, 2000). Even in contemporary times, women in STEM continue to face obstacles such as stereotype threat (Spencer, Steele, & Quinn, 1999), implicit biases in hiring and evaluation (Moss-Racusin et al., 2012), and lack of representation in leadership roles (National Science Foundation [NSF], 2021). These historical and systemic barriers collectively contribute to the persistent gender gap in STEM fields, underscoring the need for targeted interventions to dismantle structural inequities and promote greater inclusion.

Stereotype Threat and Its Impact on Girls in STEM

Research indicates that stereotype threat can hinder girls' performance in STEM subjects as early as elementary school. Spencer, Steele, and Quinn (1999) demonstrated that when girls are reminded of gender stereotypes before completing a math test, their performance suffers compared to boys. This effect is attributed to cognitive and emotional processes, including heightened anxiety, reduced working memory capacity, and self-doubt (Schmader & Johns, 2003). Similarly, Nguyen and Ryan (2008) conducted a meta-analysis showing that stereotype threat significantly affects women's performance in math-intensive tasks, particularly when stereotypes are explicitly highlighted. Stereotype threat is also linked to long-term academic and career outcomes. Beasley and Fischer (2012) found that stereotype threat contributes to the attrition of women and minorities from STEM majors in college, as it fosters a sense of exclusion and self-doubt that can lead students to disengage from these fields. When students internalize stereotypes suggesting that they are less capable in STEM due to their gender or race, they may experience heightened anxiety and reduced academic self-efficacy, making persistence in these disciplines more difficult (Steele, 1997; Walton & Spencer, 2009). Additionally, the lack of representation in STEM classrooms and faculty can further reinforce these stereotypes, creating an environment where women and minority students struggle to develop a sense of belonging (Cheryan, Plaut, Davies, & Steele, 2009). Research also suggests that stereotype threat can lead to increased cognitive load, reducing working memory capacity and overall performance, which can further undermine students' confidence and motivation to continue in STEM fields (Schmader, Johns, & Forbes, 2008). Without targeted interventions, such as mentorship programs, inclusive curriculum design, and stereotype-safe environments, the effects of stereotype threat may persist, perpetuating the cycle of underrepresentation in STEM (Dasgupta, 2011). This phenomenon is particularly concerning given the underrepresentation of women in STEM careers, which has implications for workforce diversity and innovation (Ceci & Williams, 2011). Despite the widespread acknowledgment of stereotype threat's impact, relatively little research has explored how the phenomenon interacts with other variables, such as cultural context.

STEM Interventions to Mitigate Stereotype Threat

Several interventions have been designed to reduce stereotype threat and its impact on girls in STEM. One common evidence-based intervention is values affirmation, which encourages students to reflect on their core values. Miyake et al. (2010) found that a brief values-affirmation exercise reduced the gender achievement gap in college physics courses. Similarly, Good, Aronson, and Inzlicht (2008) demonstrated that mentoring programs emphasizing beliefs that abilities can be developed through effort helped mitigate stereotype threat among middle school girls in math. Another promising strategy is increasing the representation of women in STEM. Research by Dasgupta (2011) highlights the "stereotype inoculation model," which posits that exposure to female role models in STEM can buffer against the effects of stereotype threat. This is supported by experiments showing that girls perform better on math tasks when instructed by a female teacher or mentor (Marx & Roman, 2002). Additionally, it is empirically supported that creating stereotype-safe environments, such as classrooms that emphasize collaboration and de-emphasize competition, has been shown to reduce stereotype threat (Murphy et al., 2007).

The literature reveals significant gaps that this meta-analysis aims to address. While numerous studies have examined stereotype threat in isolation, comparisons have not been made across interventions to determine which are most effective and for which outcome (s). Additionally, previous meta-analyses have often examined broad populations, with limited attention to contextual and implementation factors that may influence intervention effectiveness (Nguyen & Ryan, 2008). Additionally, this study addresses that gap by exploring specific moderating variables such as intervention type, delivery agent, grade level, and study origin. By examining these moderators, this study provides a more nuanced understanding of the conditions under which interventions are most effective for girls in STEM, offering insights that can guide the design and implementation of more targeted and impactful educational strategies.

Theoretical Framework

Social Identity Theory (SIT), developed by Tajfel and Turner (1979), provides a theoretical framework for understanding the impact of stereotype threat on girls in STEM. This theory explains how individuals derive self-esteem and identity from their membership in social groups, such as gender, and how these group memberships influence behavior and self-concept. In STEM, girls are part of a group that has historically been subjected to stereotypes suggesting they are less competent in math and science (Cheryan et al., 2017). When these stereotypes are activated in environments where gender identity is salient, such as male-dominated classrooms or workplaces, girls may experience anxiety, reduced working memory, and decreased performance due to stereotype threat (Steele & Aronson, 1995; Murphy et al., 2007). Social Identity Theory highlights how such threats are deeply rooted in systemic and group dynamics rather than being merely individual phenomena, making it a powerful tool for examining the structural inequities that perpetuate gender disparities in STEM. Historically, this theory is evident in the ways dominant social groups have maintained hierarchical structures that marginalize others. For example, during the era of racial segregation in the United States, laws and societal norms institutionalized group-based discrimination, reinforcing social hierarchies that disadvantaged African Americans while preserving white privilege (Omi & Winant, 1994). Similarly, the exclusion of women from higher education and professional fields for much of history illustrates how systemic group dynamics reinforce power imbalances. In STEM, for instance, the historical underrepresentation of women has been perpetuated by cultural narratives positioning men as naturally more competent in technical fields, leading to the entrenchment of gender disparities (Ceci, Williams, & Barnett, 2009). These patterns align with Social Identity Theory's premise that individuals tend to favor their in-group while perceiving out-groups as less capable or less deserving of opportunities, thereby sustaining structural inequalities (Sidanius & Pratto, 1999). Understanding these historical examples highlights how group-based identity processes contribute to systemic barriers, necessitating deliberate interventions to dismantle long-standing inequities. Furthermore, the theory emphasizes that individuals categorize themselves and others into in-groups and out-groups, and this process often leads to social comparisons. For girls in STEM, such comparisons may reinforce feelings of inferiority when boys, often seen as the default high-performing group, dominate STEM spaces (Nguyen & Ryan, 2008). The salience of girls' gender identity in these environments exacerbates their vulnerability to stereotype threat, particularly in competitive or evaluative contexts (Spencer, Steele, & Quinn, 1999).

Social Identity Theory also sheds light on how various social roles and affiliations shape individual experiences, allowing researchers to explore how overlapping identities, such as gender and race, compound the effects of stereotype threat. For instance, a Black girl in STEM may face unique challenges arising from both racial and gender stereotypes, a phenomenon that further complicates her experiences and underscores the need for intersectional approaches (Crenshaw, 1991; Shapiro, 2011). Historically, systemic group dynamics have reinforced these intersecting barriers. Years ago, Black women were largely excluded from scientific and mathematical fields due to both racial discrimination and gender biases, limiting their access to education and professional opportunities (Mickelson, 2003). Even after the formal end of segregation, implicit biases persisted, as seen in the underrepresentation of Black women in STEM leadership positions and academic faculties (Settles et al., 2016). These systemic obstacles align with Social Identity Theory's assertion that dominant groups maintain social hierarchies through exclusionary practices, reinforcing marginalization across multiple identity dimensions (Sidanius & Pratto, 1999). Such historical patterns illustrate how intersecting stereotypes shape group dynamics, influencing not only individual aspirations but also institutional structures that continue to perpetuate disparities in STEM fields. This framework also informs interventions aimed at mitigating stereotype threat by focusing on reshaping group dynamics and social identities. Interventions such as introducing female STEM role models can challenge the perception of STEM as an out-group and help normalize women's presence in these fields (Dasgupta, 2011). Similarly, creating inclusive classroom environments that emphasize collaboration over competition can reduce the salience of stereotypes and foster a sense of belonging (Murphy et al., 2007). Encouraging girls to reflect on their core values and strengths outside of gender stereotypes has also been shown to reduce anxiety and improve performance in STEM tasks (Good, Aronson, & Inzlicht, 2008). Despite its significant contributions, the application of SIT in stereotype threat research reveals gaps that warrant further exploration. For example, while most interventions focus on individual-level changes, systemic changes such as policy reforms and curriculum redesigns remain underexplored. Additionally, cultural differences in the salience of gender stereotypes and their impact on STEM engagement require more attention. Social Identity Theory, by emphasizing the interplay between individual psychology and social structures, not only deepens our understanding of stereotype threat but also provides actionable insights for creating more inclusive STEM environments. This makes SIT a compelling framework for advancing research and interventions aimed at closing gender gaps in STEM.

Moderating Variables

Several variables may influence the impact of stereotype threat on academic performance. Insights into these moderating factors largely stem from meta-analyses investigating stereotype threat in various contexts, including gender and racial disparities in STEM performance. This study examines how variables such as age/grade level, intervention type, delivery agent and the cultural context/origin of the study can shape the intensity and scope of stereotype threat effects, influencing whether individuals are vulnerable to or resilient against stereotype-based performance decrements. Each of these moderating variables is discussed in greater detail below.

Age or Grade Level

Previous research has established that age plays a critical role in moderating the effects of stereotype threat, with older students being more susceptible than younger ones. Nguyen and Ryan (2008) conducted a meta-analysis that demonstrated stereotype threat effects are significantly stronger among older students, particularly those in high school and college, compared to elementary-aged children. This increased susceptibility in older students can be attributed to their heightened awareness of societal stereotypes and their internalization of these stereotypes, which often amplifies cognitive and emotional interference during evaluative tasks (Nguyen & Ryan, 2008). Younger children, by contrast, may lack the developmental maturity to fully comprehend stereotypes or their implications, offering a degree of natural protection (Flavell, 1979). The developmental differences in metacognitive abilities further explain this discrepancy; adolescents and adults are more likely to engage in reflective thinking about stereotypes that apply to their social group, making them more vulnerable to their effects (Steele, 1997). These findings underscore the importance of implementing STEM interventions early in the educational trajectory when students are less likely to have fully internalized stereotypes. Early intervention can mitigate the impact of stereotype threat later in life by fostering resilience and promoting positive self-perceptions.

Study Origin/Cultural Context

The geographical and cultural context of a study, referred to here as study origin, can significantly shape how stereotype threat manifests and how effective interventions are. Stereotype threat is a socio-cultural phenomenon influenced by local norms, gender ideologies, and educational policies (Spencer et al., 2016). For instance, studies conducted in Western contexts such as the United States or Europe often address stereotype threat within individualistic cultural frames, whereas studies from Asian contexts may involve collectivist interpretations of identity and belonging (Markus & Kitayama, 1991). Cultural variability in gender roles and educational access can thus moderate the potency of stereotype threat and the applicability of specific intervention strategies. Understanding study origin as a moderator helps contextualize findings and enhances the generalizability of the meta-analysis.

Intervention Type

The type of intervention plays a crucial role in determining the success of stereotype threat reduction strategies in STEM contexts. Interventions may range from values affirmation exercises and growth mindset training to exposure to female role models and classroom-based structural changes. Each intervention type is designed to target different cognitive, emotional, or social mechanisms that underlie stereotype threat. For instance, values affirmation interventions aim to bolster self-integrity and reduce stress by prompting individuals to reflect on their core values (Cohen et al., 2006), while exposure to counter-stereotypical role models can challenge internalized beliefs about gender and STEM (Dasgupta, 2011). The effectiveness of these interventions often varies based on the psychological mechanism they engage. Meta-analytic work has shown that tailored interventions addressing specific stereotype pathways (e.g., identity threat, belonging uncertainty) tend to be more effective (Walton & Cohen, 2011). Therefore, categorizing intervention types is essential for understanding what works best in reducing stereotype threat among girls in STEM.

Delivery Agent

Who delivers the intervention whether it is a teacher, researcher, peer, or automated platform can significantly influence the intervention's impact. Research has shown that interventions are more effective when delivered by trusted figures who share identity-relevant traits with the participants, such as gender or racial background (Marx & Roman, 2002). For example, female teachers or mentors in STEM have been found to serve as powerful role models, enhancing girls’ engagement and reducing stereotype threat (Master et al., 2016).

Additionally, teacher-led interventions often benefit from relational trust and contextual relevance, while researcher-led interventions may offer more standardized implementation. The delivery agent thus acts not only as a conduit for the intervention but also as a moderator of its psychological reception and credibility.

The Current Study

Girls' participation and performance in STEM fields remain disproportionately low due to pervasive social, systemic, and cultural barriers, including stereotype threat. Addressing these disparities is both a moral and practical imperative, as it aligns with goals of equity, diversity, and inclusion in education and the workforce (Spencer et al., 2016; Walton & Cohen, 2007). Despite extensive research on stereotype threat, existing meta-analyses have largely focused on isolated interventions or specific demographic groups, leaving gaps in our understanding of how these interventions function across diverse contexts and populations. Furthermore, while strategies such as role model exposure, and the creation of stereotype-safe environments among others have demonstrated promise, their comparative effectiveness remains underexplored (Nguyen & Ryan, 2008; Walton et al., 2015).

This study bridges these gaps by analyzing findings from recent research on interventions for girls in math and science. Specifically, the study evaluates the overall effectiveness of these interventions and explores whether their efficacy is moderated by factors such as age/grade level, cultural context, intervention type and delivery agent (Schmader et al., 2008; Pennington et al., 2018). Understanding these dynamics will provide critical insights into how interventions can be tailored to maximize their impact, particularly for girls from marginalized or underrepresented backgrounds. The research questions that guide this study are as follows: (1) What types of interventions are identified in the literature, and how effective are they in increasing STEM knowledge and interest, fostering STEM identity and self-esteem, and changing negative perception about STEM? (2) Do moderating variables such as age/grade, cultural context/ origin of study, intervention type and delivery agent influence the effectiveness of these interventions? By addressing these questions, this study seeks to advance both theoretical and practical knowledge, offering actionable recommendations for educators, policymakers, and researchers committed to fostering equity in STEM education.

Outcome Variables

Outcome variables are the specific constructs measured to determine the effectiveness of an intervention (Borenstein et al., 2009; Lipsey & Wilson, 2001). In this meta-analysis, they reflect the academic, psychological, and attitudinal impacts of interventions on girls in STEM. These variables are essential for assessing whether interventions achieve desired goals like improving performance, interest, or identity in STEM fields (Schmader et al., 2008; Walton & Cohen, 2011). Their selection helps clarify how stereotype threat operates and how it can be mitigated (Spencer et al., 2016). The outcome variables examined in this study include STEM knowledge, STEM interest, STEM identity, self-esteem, and perception change; each representing a key domain in the effort to promote gender equity in STEM education (Eccles, 2009; Master et al., 2016). A detailed explanation of each outcome variable is provided below.

STEM Knowledge

STEM knowledge refers to an individual's conceptual understanding and mastery of subject-specific content in science, technology, engineering, and mathematics (National Research Council, 2011; Tai et al., 2006). It is typically assessed through academic performance measures, standardized tests, or task-based assessments (National Research Council, 2011). Acquiring foundational STEM knowledge is essential for long-term academic and career success in these fields (Tai et al., 2006). Research indicates that stereotype threat can impair performance on cognitive tasks and assessments, especially among girls and underrepresented minorities in STEM, by increasing stress and anxiety (Nguyen & Ryan, 2008; Spencer, Steele, & Quinn, 1999). Therefore, increases in STEM knowledge following interventions may indicate a reduction in cognitive load and performance anxiety (Schmader & Johns, 2003). In the reviewed studies, Knowledge of STEM was commonly measured through standardized math assessments or subject-specific exams. For example, Bancroft (2017) and Grant (2023) used the PSAT Math and STAAR Algebra I exams, while Song (2022) evaluated knowledge using pre- and post-intervention math test scores. Similarly, Zhao (2018) relied on recent school-based math exams to assess academic performance. In contrast, Ball (2014) employed the Basic Achievement Skills Inventory Math Skills Survey. Other studies used custom or experimental assessments tailored to their interventions. Master (2014) used a custom computer-based assessment, and Master (2017) employed a technology motivation and learning outcomes task. Gladstone (2024) assessed knowledge through a computer-based STEM task, and Smeding (2013) incorporated standardized math and verbal tests as part of a broader composite score.

STEM Interest

STEM interest encompasses a learner's curiosity, intrinsic motivation, and sustained engagement with STEM subjects, activities, or career paths. It is a robust predictor of both short- and long-term persistence in STEM domains (Eccles & Wigfield, 2002; Harackiewicz et al., 2008). However, girls often experience a decline in STEM interest beginning in middle school, a trend attributed in part to pervasive gender stereotypes and cultural messaging (Cheryan et al., 2017; Wang & Degol, 2013). Interventions that challenge these stereotypes, emphasize personal relevance, or incorporate relatable role models have shown promise in sustaining or rekindling girls' interest in STEM (Master et al., 2016; Boucher et al., 2017). In the reviewed studies, Interest in STEM was measured either behaviorally or through self-report instruments. Master (2016) used STEM course enrollment data before and after the intervention as an objective indicator of interest, while Tam (2020) utilized the internationally validated PISA 2012 STEM Interest Subscale to capture students' self-reported motivation and engagement with STEM subjects. Other studies such as Grant (2023) and Borman (2013) measured interest using surveys, though the exact instruments were not specified.

STEM Identity & Self-Esteem

STEM identity refers to how individuals perceive themselves as belonging in STEM fields and being recognized as competent participants by others (Carlone & Johnson, 2007). A well-developed STEM identity fosters persistence, resilience, and a stronger sense of belonging (Hazari et al., 2010; Chemers et al., 2011). Unfortunately, stereotype threat can disrupt identity development by conveying exclusion or reinforcing doubts about competence, especially for girls and marginalized groups (Steele, 1997; Spencer et al., 2016). Interventions that affirm self-worth, increase visibility of diverse role models, or provide affirming feedback can promote positive identity formation and deeper engagement in STEM (Dasgupta, 2011; Murphy et al., 2007). In the reviewed studies STEM identity was assessed through self-report scales that measured the extent to which students identified with STEM roles and disciplines. Gladstone (2024) employed identity-specific instruments, while Grant (2023) and Borman (2013) used general surveys designed to capture aspects of self-concept related to STEM. Smeding (2013) also included identity as part of a broader composite of self-report measures aimed at capturing students' perceptions and self-views in relation to academic contexts.

Self-esteem reflects an individual's overall sense of self-worth and perceived competence (Rosenberg, 1965; Harter, 1999). Negative stereotypes and stereotype threat experiences have been shown to erode self-esteem in academic contexts, particularly among girls and students from minoritized groups (Davies et al., 2002; Osborne, 1995). Low self-esteem can exacerbate disengagement and reduce motivation to pursue challenging tasks. Interventions that promote self-affirmation, inclusivity, or growth mindsets can buffer the effects of stereotype threat and help restore or maintain self-esteem (Cohen et al., 2006; Sherman et al., 2013). Higher self-esteem, in turn, supports academic resilience, confidence, and persistence in STEM learning environments (Chemer s& Garcia, 2001). In the reviewed studies Self-esteem was measured using self-report scales intended to assess individuals' overall sense of self-worth. Studies such as Gladstone (2024) and Smeding (2013) utilized explicit self-esteem scales, whereas Grant (2023) and Borman (2013) used more general surveys that included self-esteem as a component, likely through Likert-style items evaluating academic or personal confidence.

Perception Change

Perception change involves altering students' beliefs about STEM such as who belongs, what success looks like, and whether intelligence is fixed or malleable (Dweck, 2006). Stereotype-laden perceptions (e.g., "math is for boys") contribute to exclusion and disengagement (Good et al., 2008; Nosek et al., 2009). Interventions aimed at disrupting these perceptions through exposure to counter-stereotypical role models, inclusive messaging, or growth mindset training can lead to meaningful shifts in how students view themselves and others in STEM (Dasgupta, 2011; Cimpian et al., 2020). These cognitive and attitudinal changes often precede and support deeper changes in motivation, identity, and academic behavior (Eccles & Wigfield, 2002). In the reviewed studies, Perception change, particularly in relation to stereotypes and attitudes about STEM, was captured through various instruments focused on beliefs and implicit biases. In the reviewed studies, Zhao (2018) used a 20-item Math-Gender Stereotype Scale to measure changes in students' gender-based perceptions of math ability. Smeding (2013) employed a combination of self-report instruments designed to assess shifts in perception following the intervention. Gladstone (2024) incorporated attitudinal components into a computer-based STEM task, allowing for assessment of both performance and belief systems.

CHAPTER 2

METHODOLOGY

Search Strategy for Identifying Relevant Studies

Relevant articles for this meta-analysis were identified by searching the educational and psychological databases PsycINFO, ERIC, Google Scholar, ScienceDirect, Web of Science, and SpringerLink for peer-reviewed studies published between January 2010 and December 2024. This paper employed combinations of the descriptors “stereotype threat,” “intervention,” “girls,” and “STEM,” and examined reference lists of foundational meta-analyses (e.g., Good, Aronson, & Inzlicht, 2008; Spencer, Steele, & Quinn, 1999). A total of 140 records were retrieved and imported into Covidence to ensure a transparent, objective screening process. Covidence is a web-based collaboration software platform that streamlines the production of systematic and other literature reviews (Covidence, 2025). After removing 4 duplicate records, 136 titles and abstracts were independently screened against the paper’s inclusion criteria resulting in 111 exclusions. Full texts of the remaining 25 studies were then reviewed, and 13 were excluded for lacking control groups, insufficient effect-size data, or failing to meet design criteria, leaving 12 studies for inclusion. At each stage the paper maintained an audit trail in Covidence and applied a detailed codebook to extract sample characteristics, intervention details, outcome measures, and potential moderating variables, yielding a rigorously selected dataset for meta-analytic synthesis of effect sizes, main effect estimates, and heterogeneity.

Criteria for Inclusion of Studies

To be included in the meta-analysis, studies had to meet the following inclusion criteria: (a) involved girls in grades K-12; (b) STEM intervention group; (c) used an experimental, as these methodologies provide robust evidence for causal or correlational relationships. (d) provided the sample size, pre-test and post-test mean scores, and standard deviations for both treatment and control groups. This study prioritized peer-reviewed journal articles, published within the last 15 years because over the past 15 years, significant advancements in educational strategies aimed at reducing stereotype threat have emerged, incorporating contemporary pedagogical approaches and addressing gender biases in STEM education (Good, Aronson, & Inzlicht, 2008). Additionally, there has been a surge in funding directed toward increasing female participation in STEM, such as the recent $100 million donation to the University of Sydney to support girls in STEM (Palid et al. 2023 & The Guardian, 2025). These efforts, along with demographic shifts and increased research on STEM interventions, underscore the importance of focusing on studies from the last 15 years to capture the evolving landscape of gender equity in STEM. Also, studies for this meta-analysis had to be published in English; however, they could be published in any country. Studies' sample participants could be of all nationalities and ethnic/socio-linguistic backgrounds.

Criteria for Exclusion of Studies

Studies were excluded from the analysis for three reasons: first, any research involving post-secondary students was omitted, since our focus was exclusively on K–12 interventions for girls; secondly, qualitative studies and those lacking sufficient statistical data were removed to ensure we could compute effect sizes; and finally, studies whose interventions did not target a STEM discipline were excluded.

Data Analysis

Study Screening and Coding Procedures

Studies were screened for inclusion and exclusion criteria by the lead researcher in collaboration with a data assistant recruited from North Carolina State University Libraries. The data assistant is the Research Librarian for Business, Education, and Immersive Pedagogy at NC State. The lead researcher coded all studies and created a code book, which was discussed with the data assistant and academic advisor (see Appendix A). The codebook outlines the procedures for making decisions based on inclusion and exclusion criteria. The data assistant independently coded the studies using the codebook. There were two rounds of practice with examples. Independent coding was compared and discussed to ensure consensus between coders. Minor disagreements were resolved through further discussion and subsequent agreement (Hertlein, 2014).

Effect Size Calculation

After all eligible studies were coded, effect size estimation for this meta-analysis was conducted using Hedges' g, a standardized mean difference that adjusts for small sample bias and is particularly suitable for studies with varying group sizes and pre-post-control designs (Hedges & Olkin, 1985). To ensure consistency across studies with unequal group sizes and differing pre-post structures, the "Effect Size for Mean Differences of Groups with Unequal Sample Size within a Pre-Post-Control Design" calculator provided by Psychometrica (https://www.psychometrica.de/effect_size.html) was employed. This tool facilitated the computation of standardized effect sizes while accounting for pre-intervention differences and control group outcomes. This tool facilitated the computation of standardized effect sizes while accounting for pre-intervention differences and control group outcomes. In addition to the effect size, the variance of each Cohen’s d value was calculated using the formula proposed by Marfo and Okeyere (2019):

$$V_d = \left(\frac{n_1+n_2}{n_1 n_2}\right) + \left(\frac{d^2}{2(n_1+n_2)}\right),$$

where n1 and n2 represent the sample sizes of the treatment and control groups, and d is the computed Cohen’s d. Hedges’ g was then derived from Cohen’s d by applying a correction factor to reduce positive bias in small samples. Effects below 0.20 are typically considered small and potentially insignificant in applied psychological settings, while values above 0.80 are viewed as large and practically meaningful, offering a standardized framework for interpreting intervention impact (Cohen, 1988; Ferguson, 2009).

Meta-Analytic Model and Heterogeneity Assessment

The main effect for interventions was estimated using a random-effects meta-analysis, an approach that accommodates both within-study and between-study variance and assumes that true effect sizes differ across contexts (Borenstein et al., 2009). However, as Bryan, Tipton, and Yeager (2021) argue, it is not sufficient to focus only on average effects; behavioral science must embrace heterogeneity as informative rather than treat it as noise; the hallmark of what they describe as a heterogeneity revolution. Accordingly, this study assessed heterogeneity statistically using the Q statistic, which tests the null hypothesis of a common effect size across studies, and the I^2 statistic, which quantifies the proportion of total variability due to heterogeneity rather than chance. Significant Q ($p < .05$) and I^2 values exceeding 50% were interpreted as indicators of substantial heterogeneity, warranting moderator analyses.

Empirical work in mathematics education illustrates why this perspective is critical. In a large-scale meta-analysis of U.S. PreK-12 randomized trials (191 studies; 1,109 effect sizes), Williams, Doabler, Smolkowski, Clarke, and Baker (2022) found a modest overall effect ($g \approx 0.31$) but a wide range of impacts (-0.60 to 1.23), underscoring substantial heterogeneity in math intervention outcomes. Their moderator analyses revealed that outcome type (researcher-created vs. standardized), delivery format (technology-based vs. teacher-led), intervention duration, grade level, and publication year all accounted for meaningful variation in effects (Williams et al., 2022; Institute of Education Sciences, 2022). By foregrounding both statistical heterogeneity and its substantive sources such as intervention mode, assessment type, student age, and socioeconomic context this approach aligns with Bryan et al.'s (2021) call: rather than smoothing away variability, we must examine for whom, in what settings, and under what conditions math interventions work best. In doing so, the analysis contributes to a more nuanced and equitable understanding of intervention effectiveness that resists simplistic aggregation and instead highlights meaningful subgroup differences.

Moderator Analyses and Statistical Software

To examine the influence of moderator variables on intervention effectiveness, subgroup analyses and meta-regression techniques were employed. Moderator variables: intervention type, delivery agent, grade level, and study origin /cultural context were coded categorically based on study characteristics. Each moderator was independently tested using a random-effects model to assess whether effect sizes significantly differed across subgroups. This approach allowed for the identification of patterns in intervention outcomes relative to contextual factors, providing a deeper understanding of which conditions enhance or diminish the impact of interventions for girls in STEM. Moderator variables in this meta-analysis were coded as study-level characteristics rather than experimental manipulations. For example, intervention type (e.g., growth mindset, self-affirmation, role-model exposure), delivery agent (e.g., teacher vs. expert), grade level, and cultural context were extracted based on how the primary studies described their design and implementation. These variables were not typically manipulated within studies (e.g., no randomized comparisons of teacher- vs. expert-delivered conditions) but instead reflected naturally occurring variation across the literature. A small subset of studies reported fidelity procedures (e.g., scripted lesson plans, adherence checks), but reporting was inconsistent and often descriptive rather than quantitative (see Bancroft, 2017; Master, 2017; Zhao, 2018). Accordingly, moderator analyses should be interpreted as exploratory and descriptive, identifying patterns across contexts rather than causal tests of systematically implemented conditions (Borenstein et al., 2009).

All analyses were conducted using Comprehensive Meta-Analysis (CMA) software. CMA is a specialized statistical software used to conduct meta-analyses across a wide range of research domains. It enables users to input data in multiple formats and compute summary statistics using fixed-effect, random-effects, or mixed-effects models. In addition, CMA supports subgroup analyses, publication bias detection, and sensitivity testing, making it a widely used tool for synthesizing quantitative research findings in a rigorous and replicable manner (Borenstein, Hedges, Higgins, & Rothstein, 2013).

CHAPTER 3

RESULTS

What is the Efficacy of Interventions in Overcoming Stereotype Threat and Improving STEM Outcomes?

A total of 111 Studies were excluded from this analysis for using post-secondary participants, lacking sufficient quantitative data to calculate effect sizes, and or focusing on interventions unrelated to STEM disciplines. Table 1.5 presents studies that met criteria for full text review but did not meet the full inclusion criteria. Twelve experimental and quasi-experimental studies met inclusion criteria for this meta-analysis, representing more than three thousand school-aged girls (Grades 1–12) across diverse educational settings in the United States, Europe, and China. Table 1.1 provides a description of studies included and interventions used. These interventions fell into several broad categories, including cognitive-reframing approaches (e.g., growth-mindset training, non-stereotypical classroom structures), self-affirmation exercises, hands-on technology experiences, female-teacher contact, role-model exposures, and online workshops, and were delivered either by STEM specialists ("expert-directed") or by regular classroom teachers. Across eighteen distinct outcome-measure pairings, effect sizes (Hedges' g) ranged from negligible ($g = 0.02$) to large ($g = -0.92$), with an overall weighted mean of $g = 0.18$ (95% CI [0.12, 0.24], $p < .001$). A Q-test for heterogeneity confirmed substantial variation among studies ($Q = 48.2$, $df = 11$, $p < .001$; $I^2 = 77.2\%$), indicating that moderator analyses were warranted to explain these differences.

In eight out of the twelve studies included in this meta-analysis, researchers conducted pre-intervention surveys to verify the presence of stereotype threat among participating girls. In the remaining four studies, participants were assigned to control and experimental groups without a pre-intervention survey to establish baseline levels of stereotype threat. The remaining 4 studies assigned participants to the control and experimental group without a pre-survey to confirm the presence of stereotype threat. These baseline assessments were crucial in establishing the relevance and necessity of targeted interventions. To measure stereotype threat, most studies employed validated self-report instruments such as the Stereotype Vulnerability Scale (Steele, 1997) or adapted versions of the Stereotype Threat Scale developed by Spencer et al. (1999). These tools assessed participants' awareness of gender-based academic stereotypes, their perceived susceptibility to those stereotypes, and the extent to which such perceptions influenced their confidence and academic self-concept in STEM domains. By confirming the psychological salience of stereotype threat prior to intervention, these studies ensured that the strategies employed were contextually appropriate and targeted toward students genuinely at risk of stereotype-driven underperformance. The results of this meta-analysis (Table 1.1, 1.2, 1.3 and 1.4) are organized according to four key outcome domains: STEM knowledge, STEM interest, STEM identity and self-esteem, and perception change. Table 1.2 presents the outcomes, measurement instruments, and corresponding effect sizes reported in each study.

STEM Knowledge

Seven studies evaluated knowledge gains and advancement through standardized exams (e.g., PSAT Math, STAAR Algebra I) or researcher-designed quizzes. The study with the highest effect size was Master's (2017) first-grade intervention which was a brief, hands-on technology experience that yielded a very large effect ($g = 0.83$) on a technology-motivation survey. Bancroft (2017) reported a moderate effect ($g = 0.42$) on high-stakes math tests following a brief, expert-led self-affirmation exercise in Grade 9, while Master (2014) observed a smaller knowledge gain ($g = 0.15$) on a computer-based STEM test after increasing female-teacher contact in Grades 10–12. In China, Song (2022) and Zhao (2018) each documented small improvements ($g = 0.14$ and 0.08, respectively) on in-class math exams following growth-mindset and multifaceted self-affirmation programs. Ball (2014) found only a negligible effect ($g = 0.05$) for a gender-fair curriculum in Grade 9, and Smeding (2013) recorded a small gain ($g = 0.08$) on math performance when reversing the order of math and verbal tests for Grade 8 girls. A random-effects synthesis of these seven knowledge outcomes yielded a pooled effect size of $g = 0.22$ (95% CI [0.10, 0.34], $p < .001$) but also substantial heterogeneity ($Q = 18.7$, $df = 6$, $p = .005$; $I^2 = 68\%$), suggesting that age, intervention design, and assessment type meaningfully influenced knowledge gains.

STEM Interest

Six studies measured changes in girls' interest in STEM through self-report scales, enrollment intentions, or international assessments. Master (2016) observed a large effect ($g = 0.61$) on course-enrollment intentions after restructuring the classroom environment to reduce gendered cues for Grades 10–12. Grant (2023) found a moderate interest boost ($g = 0.57$) in PSAT and STAAR math performance following an all-girls technology camp for Grades 5–8, though her parallel self-esteem gains were minimal ($g = 0.02$). In contrast, Tam (2020) documented a modest improvement ($g = 0.14$) on PISA ICT interest scores after an online video-making workshop for Grades 7–8, and Borman (2013) reported a moderate effect ($g = 0.20$) on multiple achievement measures after affirmations in Grade 7. European cohorts showed smaller gains: Smeding (2013) reported $g = 0.07$ for STEM identity and $g = 0.094$ for self-esteem tied to interest, and Gladstone (2024) found $g = 0.06$ on computer-test interest in Grades 1–3. When these six interest effect sizes were synthesized, the pooled estimate was $g \approx 0.27$ (95% CI

[0.13, 0.41], $p < .001$), with moderate heterogeneity ($Q = 9.8$, $df = 5$, $p = .08$; $I^2 = 49\%$), signaling reasonably consistent positive shifts in interest across varied contexts.

STEM Identity and Self-Esteem

Eight studies explicitly targeted girls' STEM identity and general academic self-concept. Borman (2013) and Smeding (2013) reported small gains in self-esteem ($g \approx 0.094$) after self-affirmation exercises. Zhao (2018) found a modest identity boost ($g = 0.08$) post intervention in Grades 7–8. Gladstone's (2024) early elementary role-model exposure yielded stronger effects; $g = 0.35$ on identity and $g = 0.56$ on self-esteem; underscoring young learners' responsiveness to visible exemplars. Grant (2023) saw minimal self-esteem change ($g = 0.02$) despite moderate interest gains, indicating that knowledge and interest interventions do not necessarily translate into identity shifts. Aggregating these eight measures produced a pooled $g = 0.19$ (95% CI [0.08, 0.30], $p = .001$), with moderate heterogeneity ($Q = 12.4$, $df = 7$, $p = .09$; $I^2 = 44\%$). Thus, while interventions generally enhanced identity and self-esteem, effect magnitudes varied by approach and age.

Change in Stereotypical Perception about STEM

Among the studies included in this meta-analysis, only one provided quantifiable post-intervention data specifically related to changes in gender-stereotypical perceptions about STEM. Zhao et al. (2018) employed a 20-item math-gender stereotype scale, which separately measured beliefs about boys' and girls' math abilities (e.g., "boys are good at math problem solving" and "girls are good at math problem solving"). The intervention led to a standardized mean difference (SMD) of $g = -0.92$, indicating a significant reduction in stereotype endorsement among participants in the intervention group compared to those in the control group. The main effect was statistically significant, $F(1, 74) = 3.84$, $p = .05$, with a partial eta squared (η^2) = 0.05, suggesting a small to moderate effect. Heterogeneity analysis revealed a moderate level of variation across the data, $F(1, 73) = 6.49$, $p = .01$, with partial $\eta^2 = 0.08$. Other studies included in the review (e.g., Ball, 2014; Master, 2014) discussed perception-oriented changes narratively or incorporated them within broader outcome constructs but did not provide the statistical data.

Does the Effectiveness of STEM Interventions for Girls Vary Based on Grade Level, Intervention Type, Delivery Agent, or Origin of Study

Moderator Analysis

Given the high degree of variability across these twelve studies ($I^2 = 77.2\%$), this paper explored several moderators, including intervention type, delivery agent, grade level, study origin, and the cultural context of participants, to better understand which factors amplify or attenuate girls' gains in STEM.

Intervention Type. Cognitive-reframing approaches, particularly growth-mindset trainings conducted in China (Song, 2022) and the United States (Master, 2017), produced the largest average effects ($g \approx 0.34$), significantly above the overall mean ($\beta = .12$, $SE = .05$, $p = .02$). Self-affirmation exercises delivered in the U.S. (Bancroft, 2017) and China (Zhao, 2018) as well as in Switzerland (Smeding,e 2013) yielded smaller but reliable gains ($g \approx 0.16$). Early elementary role-model exposures (Gladstone, 2024, U.S.) and brief hands-on technology experiences (Master, 2017, U.S.) drove some of the largest single-study effects ($g = 0.83$) on both knowledge and identity, suggesting that concrete, developmentally appropriate engagements can be especially powerful.

Delivery Agent. Interventions led by STEM specialists (expert-directed) across U.S., European, and Chinese contexts averaged $g = 0.23$, slightly outperforming those implemented by regular classroom teachers ($g = 0.17$). Although a meta-regression indicated this difference did not reach statistical significance ($\beta = .06$, $SE = .04$, $p = .12$), qualitative examination suggests that expert-led workshops often include more consistent fidelity checks and specialized materials, whereas teacher-led programs may vary more by individual educator comfort and training.

Grade Level. Developmental timing emerged as a robust moderator. Early-elementary interventions (Grades 1–3; all U.S. studies) produced the largest pooled gains ($g = 0.38$; 95% CI [0.22, 0.54]), particularly in predominantly white and mixed-ethnicity classrooms. Middle-school programs (Grades 5–8; U.S., China, France) yielded moderate gains ($g = 0.21$; 95% CI [0.10, 0.32]), while high-school interventions (Grades 7–12; U.S., China) showed the smallest improvements ($g = 0.15$; 95% CI [0.05, 0.25]). An ANOVA confirmed this developmental trend ($F(2, 9) = 4.12$, $p = .04$), underscoring the outsized impact of early, preventive engagement.

Study Origin & Ethnicity. Although formal comparisons by region and participant ethnicity were limited by incomplete demographic reporting in most studies, some patterns emerged descriptively. U.S. studies of predominantly Hispanic or African American cohorts (e.g., Bancroft 2017; Borman 2013) reported average effects ($g \approx 0.18$) similar to those in majority-white samples (Ball 2014; Gladstone 2024). Chinese interventions (Song 2022; Zhao 2018; Tam 2020) likewise achieved effects ($g \approx 0.16$) on par with Western contexts, suggesting broad cross-cultural applicability. European work (Smeding 2013) with majority-Caucasian samples yielded somewhat lower gains ($g \approx 0.08$) in the absence of complementary growth-mindset framing, perhaps reflecting differences in curricular emphasis or cultural norms around gender and math.

Durability & Publication Bias

Limited follow-up data preclude strong conclusions about long-term persistence. Bancroft (2017) documented sustained math gains six months post-intervention (Δ_g_ = 0.03), and Ball (2014) reported negligible decay in BASI scores over a semester ($\Delta\_g\_ = 0.01$). Gladstone (2024) did not report follow-up outcomes. Funnel-plot asymmetry (Egger's intercept = 2.11, $p = .07$) and trim-and-fill adjustments adding two imputed null effects reduced the overall g to 0.16. Leave-one-out sensitivity analyses showed no single study unduly influenced the pooled estimate (range $g = 0.17–0.20$).

CHAPTER 4

DISCUSSION

This meta-analysis reveals a consistent pattern of small to moderate benefits from school-based interventions designed to improve girls' STEM outcomes, including knowledge ($g = 0.22$), interest ($g = 0.27$), and identity/self-esteem ($g = 0.19$). Although these effect sizes may appear modest, they are meaningful when contextualized within the broader educational literature, where even small improvements across large populations can yield substantial cumulative impacts over time (Lipsey et al., 2012; Kraft, 2020). This principle is especially relevant in educational interventions, where small average gains at the individual level can scale into large societal benefits when interventions are applied systemically. As Kraft (2020) argues, even interventions with effect sizes in the range of 0.10 to 0.30 can translate into months of additional learning if sustained across grade levels and cohorts. Lipsey et al. (2012) similarly highlights that small standardized mean differences in early education can lead to meaningful downstream effects, particularly in reducing long-term achievement gaps. When interventions reach thousands or potentially millions of students over time, even small shifts in academic achievement, motivation, or identity can reshape educational trajectories, influence course-taking patterns, and ultimately increase representation in STEM careers. Therefore, modest effect sizes in the context of a scalable, evidence-based approach should not be dismissed, particularly when targeting historically underrepresented groups where equity gains have multiplicative social and economic consequences. Importantly, the data point toward several highly effective strategies, particularly growth-mindset training, early role-model exposure, and hands-on learning experiences that warrant closer examination for both policy and practice.

Growth Mindset Interventions: A Cornerstone for Cognitive Reframing

Among the various intervention types analyzed, growth-mindset training emerged as the most robust and consistent approach across contexts. Grounded in the work of Dweck (2006), growth mindset interventions aim to reshape students' beliefs about the malleability of intelligence, thereby countering fixed, ability-based stereotypes often disproportionately internalized by girls in STEM fields (Good, Aronson, & Harder, 2008). In the current synthesis, these interventions yielded above-average effects, particularly when implemented with fidelity in structured, developmentally appropriate formats.

For example, Song (2022) in China and Master (2017) in the U.S. both demonstrated that brief yet targeted messages emphasizing effort and improvement rather than innate talent-led to significant boosts in STEM achievement and interest. These findings align with prior large-scale trials, such as Yeager et al. (2019), who showed that growth-mindset interventions, especially when well-integrated into classroom culture, can generate long-term academic benefits for underrepresented students. Moreover, growth mindset strategies can be especially potent when combined with classroom practices that reinforce competence through iterative, feedback-rich tasks; a key mechanism for promoting mastery orientation (Rattan et al., 2015). Importantly, however, growth mindset interventions are not uniformly effective in all situations; their impact is context-dependent (Yeager & Dweck, 2020; Paunesku et al., 2015). Large-scale studies and preregistered replications have shown that growth mindset effects are reliable but also meaningfully heterogeneous, with variations linked to differences in student populations, classroom climates, and implementation quality (Yeager et al., 2019; Sisk et al., 2018).

Although the concept of a growth mindset, defined as the belief that intellectual ability can be developed through effort and learning (Dweck, 2006), has received significant attention in educational psychology, its measurement and practical impact remain points of contention. Several researchers have questioned whether growth mindset interventions consistently lead to academic improvement, or whether observed effects are a product of methodological limitations and contextual variability (Sisk et al., 2018; Li & Bates, 2019). Concerns have also been raised regarding the validity of instruments used to assess mindset, particularly due to the potential for social desirability bias and the construct's sensitivity to situational factors (Credé, 2018). At the same time, recent work has clarified that while growth mindset effects are meaningfully heterogeneous across individuals and contexts, high-quality, preregistered replications and large-scale studies provide justification for confidence in growth mindset research (Yeager & Dweck, 2020). This suggests that rather than being dismissed outright, mindset interventions should be understood as context-dependent tools whose success hinges on alignment with the learners, the domain of application, and the quality of implementation. This context sensitivity helps explain both their power and their limitations. Growth mindset interventions appear to work best in environments where students are facing challenge or stereotype threat, since the message that "ability can be developed" directly counters the fixed narratives that undermine persistence and achievement (Good et al., 2003; Aronson, Fried, & Good, 2002; Yeager & Dweck, 2020). The effectiveness of these interventions also lies in their dual function: they not only bolster individual resilience but also signal to students that their academic environments value learning and growth (Blackwell, Trzesniewski, & Dweck, 2007; Yeager et al., 2014). When these two levels align students' beliefs and classroom norms growth mindset effects become especially robust (Yeager & Walton, 2011; Walton & Cohen, 2011).

Despite these critiques, the current meta-analysis found consistent evidence of improvement in math achievement following growth mindset intervention. Both within-group pre-post comparisons and between-group comparisons involving control and intervention conditions revealed statistically significant gains in math performance. These findings align with prior research indicating that growth mindset interventions are more effective when they are specifically tailored to academic content areas, particularly in STEM-related domains (Paunesku et al., 2015; Yeager et al., 2019).

The strength of this outcome may be attributed to the design of the intervention. In the study reviewed, growth mindset strategies were explicitly integrated into math instruction, often emphasizing effort, learning from mistakes, and persistence when facing mathematical challenges. This finding aligns with theoretical work suggesting that growth mindset (or incremental theory) is positively associated with self-regulatory processes such as setting learning goals (Burnette et al., 2013), the findings from this review suggest that carefully structured, content-specific interventions can produce meaningful improvements in students' math achievement. Taken together, this body of research suggests that growth mindset interventions are highly effective not because they offer a universal "one-size-fits-all" solution, but because they interact with students' immediate educational and social contexts to reshape how they interpret challenge, ability, and belonging (Yeager et al., 2019; Walton & Yeager, 2020).

The Power of Early Intervention: Developmental Timing as a Moderator

One of the most striking patterns to emerge from the moderator analysis is the pronounced impact of early intervention. Interventions administered in the early elementary years (Grades 1–3) demonstrated the largest pooled effects ($g = 0.38$), far exceeding those in middle school ($g = 0.21$) and high school ($g = 0.15$). These findings are consistent with developmental psychology research suggesting that children begin forming gendered academic beliefs as early as preschool (Cvencek et al., 2011) and that early experiences play a foundational role in shaping long-term academic identity (Eccles, 2009).

Hands-on technology activities and exposure to female STEM role models appear especially effective for younger learners, as demonstrated in Gladstone (2024) and Master (2017). Such interventions likely leverage the cognitive flexibility and openness to new experiences that characterize early childhood (Piaget, 1971), while also counteracting prevailing media and societal stereotypes that may otherwise shape self-concept. Given this developmental sensitivity, the data strongly support front-loading STEM engagement in the early years, before girls begin internalizing fixed beliefs about ability and belonging. However, as these findings are drawn from only two studies, further research is needed to confirm and expand upon this promising approach. Additionally, there remains a critical gap in longitudinal research confirming the durability of these effects. Although broader educational literature suggests early interventions can yield long-lasting academic and psychosocial benefits (Reynolds et al., 2001; Reynolds et al., 2018), few studies have tracked the long-term impact of early STEM-specific interventions for girls. Programs like the Chicago Child-Parent Center (CPC) and the REDI intervention have demonstrated that early school-based enrichment can lead to sustained improvements in academic performance and life outcomes into adolescence and adulthood (Reynolds et al., 2011; Bierman et al., 2021). However, STEM-focused initiatives such as QCamp and Camp Reach have only begun to explore these trajectories, often relying on small samples or short-term follow-ups (Milto et al., 2016). Given the promising short-term outcomes observed in studies like Gladstone (2024) and Master (2017), future longitudinal research is needed to determine whether these early gains in STEM knowledge, identity, and interest are maintained over time and translate into long-term academic persistence or career engagement.

Role Models and Representation: Building Identity Through Visibility

Exposure to relatable role models, particularly female professionals working in STEM was associated with significant gains in STEM identity and self-esteem. This finding echoes a growing body of literature underscoring the importance of representation in combating stereotype threat and enhancing a sense of belonging (Dasgupta, 2011; Lockwood, 2006). Young girls are more likely to see STEM as attainable and aligned with their own identities when they encounter diverse role models who defy traditional gender norms.

Gladstone (2024), for example, reported effect sizes as high as $g = 0.56$ in self-esteem among early elementary girls following role-model exposure, reinforcing similar findings from informal learning contexts such as STEM summer camps (Herrmann et al., 2016). Crucially, role models must be perceived as both competent and relatable; too great a social distance can undermine identification and reduce efficacy (Cheryan et al., 2011). Schools should therefore invest in bringing local women in STEM into the classroom and diversifying the materials, stories, and visuals students encounter daily.

Self-Affirmation and Psychological Safety

While not as impactful as growth-mindset or role-model interventions, self-affirmation strategies, brief writing or reflection activities aimed at reinforcing personal values nonetheless yielded small but reliable gains in identity and self-esteem. Rooted in the social-psychological literature (Steele, 1997), self-affirmation works by buffering students against identity threats, allowing them to better engage with challenging academic material.

Studies such as Borman (2013) and Zhao (2018) support this mechanism, suggesting that self-affirmation may be particularly effective in high-stakes environments where stereotype threat is activated. Moreover, because these exercises are brief, inexpensive, and easily embedded into existing curricula, they represent a scalable option for schools seeking to complement more intensive interventions. Although the effect sizes in this category were smaller ($g \approx 0.08–0.16$), the low implementation cost makes them a valuable component of a broader intervention strategy.

Who Delivers the Intervention: The Role of Experts vs. Teachers

Interventions delivered by STEM experts consistently outperformed those implemented by general classroom teachers ($g = 0.23$ vs. 0.17), though the difference was not statistically significant. This suggests that content knowledge and fidelity may modestly enhance outcomes. However, it also underscores the potential scalability of teacher-led models, provided they are supported with structured materials, training, and implementation guidelines. Teacher-led interventions can be especially powerful when embedded within a supportive school culture that values equity, inquiry, and innovation (Darling-Hammond et al., 2020). Investing in teacher professional development is, therefore, a critical step toward ensuring broader adoption and sustainability of effective STEM interventions.

The Need for Long-Term Follow-Up and Systemic Integration

One notable limitation across the included studies is the lack of long-term follow-up data. Only a few studies, such as Bancroft (2017), examined persistence of gains over time, and even these were limited to 4–6 months. Without longitudinal data, it is difficult to assess whether short-term boosts in interest or self-esteem translate into lasting changes in academic choices, course enrollment, or career aspirations. Longitudinal research is essential to evaluating the extent to which early interventions shift trajectories and narrow the gender gap in real-world STEM outcomes (Tai et al., 2006).

Furthermore, while isolated interventions offer proof of concept, broader systemic change will require integration of these strategies into schoolwide and district-wide policies. For example, embedding growth-mindset principles into teacher training, curriculum design, and assessment practices could amplify their impact. Similarly, partnerships with STEM professionals and community organizations can help sustain role-model exposure beyond isolated events.

A Multi-Tiered, Developmentally Informed Model

Taken together, these findings support a multi-tiered, developmentally sensitive model for advancing girls' participation and success in STEM. In early childhood (Grades 1–3), schools should prioritize hands-on STEM exploration, growth-mindset language, and role-model engagement. In middle school (Grades 5–8), efforts should shift to include cognitive reframing, classroom environment restructuring, and ongoing identity affirmation. In high school, where gender disparities in advanced math and science track placement widen (Wang & Degol, 2017), targeted growth-mindset modules and teacher training become essential. Such a layered approach recognizes that no single intervention is a panacea. Instead, sustained, culturally responsive programming across the K–12 continuum is needed to meaningfully disrupt the patterns of exclusion and underrepresentation that have long characterized STEM fields.

Cultural and Racial Analysis

The findings of this study must also be understood within broader cultural and racial contexts. For instance, a Black girl pursuing STEM is uniquely situated at the intersection of both racial and gender stereotypes, which can amplify barriers to belonging and achievement. This intersectional perspective, first articulated by Crenshaw (1991), highlights that overlapping systems of disadvantage create experiences that cannot be reduced to race or gender alone. In STEM education, such students may simultaneously navigate the stereotype that "girls are not good at math" alongside racialized expectations that question the intellectual capacities of Black learners, producing a compounded psychological burden that can undermine engagement, persistence, and performance (Shapiro, 2011).

Most studies of stereotype threat and growth mindset interventions have been conducted in Western contexts, such as the United States or Europe, where identity is often conceptualized within individualistic cultural frames. In these contexts, interventions typically emphasize personal agency, self-regulation, and individual resilience. However, as Markus and Kitayama (1991) argue, individuals in collectivist cultures such as many Asian societies construct identity in relation to social groups and interdependent belonging. In such contexts, stereotype threat may manifest less as fear of personal failure and more as anxiety about letting down one's group, while growth mindset interventions may be more effective when they are framed in terms of collective progress or family honor rather than individual achievement. This suggests that cultural variations shape not only the experience of stereotype threat but also the efficacy of interventions intended to counteract it.

A racial analysis further underscores the systemic forces that complicate identity development in STEM. Beyond individual stereotypes, structural inequities such as unequal access to advanced coursework, underrepresentation of teachers of color, and the marginalization of culturally relevant pedagogies continue to limit opportunities for Black and Brown students (Nasir et al., 2020). These institutional patterns reinforce stereotype threat by embedding racialized messages into the educational environment, thereby shaping both expectations and outcomes. Thus, interventions that focus solely on individual cognition or mindset without addressing systemic inequities risk overlooking the broader cultural and racial dynamics that perpetuate disparities.

Taken together, these insights point to the necessity of adopting intersectional and culturally responsive approaches when designing and implementing interventions in STEM education. Addressing stereotype threat requires not only fostering individual growth mindsets and providing role models but also transforming classroom environments and institutional practices in ways that validate diverse identities and dismantle systemic barriers.

Interpreting Intervention Effects Through the Lens of Social Identity Theory

The findings of this study can be explained through the lens of Social Identity Theory, which posits that individuals derive a significant portion of their self-concept from their membership in social groups and the value they attach to those groups (Tajfel & Turner, 1979). In the context of STEM education for girls, interventions that enhance identification with the STEM domain through exposure to female role models, inclusive classroom cues, and opportunities for meaningful engagement serve to shift the perception of STEM as a socially distant or male-dominated space (Cheryan et al., 2015; Dasgupta, 2011). The observed improvements in girls' STEM interest, knowledge, and self-concept, particularly among younger age groups, align with the theory's assertion that positive social comparisons and ingroup representation foster stronger identification and motivation (Tajfel & Turner, 1986; Walton & Cohen, 2011). For instance, early exposure to relatable role models not only provided cognitive scaffolding but also helped girls reframe STEM as a domain to which they can belong (Betz & Sekaquaptewa, 2012; Lockwood, 2006). Similarly, growth mindset training and self-affirmation activities contributed to a more positive academic identity, supporting the idea that internalizing positive group-relevant attributes (e.g., "girls can succeed in STEM") promotes engagement and persistence (Dweck, 2006; Yeager & Dweck, 2020). The diminishing effects among older girls may reflect more entrenched perceptions of group boundaries and stereotypes, suggesting that once a social identity is shaped by exclusion or negative comparison, it becomes more resistant to change (Steele, 1997; Eccles, 2009). Thus, the efficacy of early interventions can be understood as a function of timing within the identity development process: the earlier girls encounter affirming experiences in STEM, the more likely they are to internalize a STEM identity and resist stereotype-based disengagement (Master et al., 2016; Murphy et al., 2007). This underscores the importance of not only fostering STEM competence but also reinforcing group belonging and self-relevance within the educational environment (Good et al., 2003; Dasgupta & Stout, 2014).

Measures Used and Psychometric Considerations

Studies included in this meta-analysis employed a range of instruments to assess math achievement and related outcomes, from well-established standardized tests to researcher-developed or affective measures with limited psychometric documentation. For example, Ball (2014) used the Basic Achievement Skills Inventory - Math Skills Survey, a tool with limited publicly reported validation data, raising questions about construct precision (Li & Bates, 2019). Similarly, Song (2022) relied on pre- and post-intervention math test scores, though without reference to standardized or normed instruments, potentially limiting interpretability. Zhao (2018) used recent school-based math assessments alongside a custom 20-item Math-Gender Stereotype Scale, both of which lacked reported reliability or validation procedures. In contrast, studies such as Borman (2013) and Grant (2023) utilized validated standardized measures like the MAP Spring Assessments, WKCE, STAAR Algebra I, and PSAT Math, which offer stronger psychometric foundations and curricular alignment (Ferguson, 2009; NWEA, 2020). Tam (2020) employed the PISA 2012 STEM Interest Subscale, which, while internationally validated, has been critiqued for cultural generalizability and alignment with national curricula (OECD, 2014).

Studies like Smeding (2013) and Gladstone (2024) relied heavily on self-report scales for constructs like STEM identity, self-esteem, and perception change, though most failed to report internal consistency or factor structure, limiting construct validity (Credé, 2018). Master (2017) used a customized Technology Motivation & Learning Outcomes Task- a non-standardized measure developed for a single experimental context which, despite its contextual relevance, lacked any psychometric reporting. Overall, the diversity in measurement quality, particularly among affective and researcher-developed tools, likely contributed to the heterogeneity observed in effect sizes. Prior reviews caution that instruments lacking in psychometric rigor may inflate or mask true intervention effects (Hyde, 2005; Sisk et al., 2018), highlighting the continued need for curriculum-aligned, validated tools in STEM education research.

Recommendations for Practice: Enhancing Girls' Engagement and Achievement in STEM

Based on the reviewed evidence, effective strategies to promote girls' participation and success in STEM should begin early in a student's academic career and evolve to meet developmental needs across time. Interventions introduced in the early elementary grades, particularly between Grades 1 and 3, are especially impactful, as gender-based stereotypes around STEM ability tend to solidify as children grow older (Cvencek et al., 2011; Tai et al., 2006). At this foundational stage, educators should embed hands-on, play-based STEM activities into the core curriculum. These experiences promote engagement, foster problem-solving skills, and build early competence, which is essential for shaping positive academic identities (Eccles, 2009; Wang & Degol, 2017). Moreover, classroom language and materials must be free from gendered assumptions. Teachers should avoid subtle messages that reinforce stereotypes, and instead, utilize texts, imagery, and examples that portray STEM as inclusive and accessible to all genders (Cheryan et al., 2011).

Early exposure to diverse and relatable female role models in STEM significantly enhances young girls' sense of belonging and possibility in these fields. These role models should be integrated into classroom instruction not only through guest speakers and mentorship programs but also through curricular content and media (Dasgupta, 2011; Herrmann et al., 2016). Exposure to "ingroup" experts, individuals with whom students identify socially or culturally, has been shown to improve performance and persistence by acting as a psychological buffer against stereotype threat (Dasgupta, 2011; Lockwood, 2006). This is particularly effective when role models reflect the students' racial, cultural, or socioeconomic backgrounds, further reinforcing identity alignment and future aspirations (Herrmann et al., 2016).

Cognitive reframing techniques, particularly growth mindset interventions, have been identified as one of the most consistent and impactful approaches across multiple domains of STEM outcomes. These interventions teach students that intelligence and ability in STEM are malleable and can be developed through effort and persistence rather than being fixed traits (Dweck, 2006; Rattan et al., 2015). When educators adopt and communicate growth-oriented beliefs, students demonstrate greater resilience, motivation, and academic performance in STEM. However, when teachers themselves convey entity theories, beliefs that some students are innately less capable, girls are particularly susceptible to internalizing these views, leading to decreased confidence and interest (Rattan et al., 2015). Thus, professional development efforts should equip teachers with strategies to promote growth-oriented feedback and normalize struggle as part of the learning process (Darling-Hammond et al., 2020).

In addition to mindset training, identity-supportive practices such as self-affirmation exercises offer valuable benefits. While the effect sizes for these interventions may be smaller, they reliably enhance self-esteem and buffer against stereotype threat, especially when integrated into everyday classroom practices (Steele, 1997; Good et al., 2008). Encouraging students to reflect on personal values, goals, and sources of pride fosters a more resilient academic identity. These practices are particularly effective when paired with inclusive classroom cues that signal belonging, such as gender-neutral décor, balanced groupings, and equitable teacher attention (Cheryan et al., 2011). When the learning environment consistently reinforces students' sense of belonging, they are more likely to view STEM as part of their possible selves (Eccles, 2009).

Though interventions led by STEM experts tend to yield slightly stronger outcomes, classroom teachers can achieve comparable results when supported with high-quality training and structured materials (Kraft, 2020). This underscores the importance of investing in sustained professional development that emphasizes not only content knowledge but also the social-psychological dimensions of teaching (Darling-Hammond et al., 2020). Teachers should be prepared to identify and disrupt microaggressions, unconscious bias, and stereotype threats in the classroom. Moreover, they should be equipped with research-based instructional strategies to build inclusive learning communities and support diverse learners.

As students progress to middle and high school, continued reinforcement is essential to sustain earlier gains. During adolescence, the influence of peer dynamics and identity salience increases, making this a critical period for intervention (Eccles, 2009; Wang & Degol, 2017). Schools should maintain a strong emphasis on growth mindset instruction, while also expanding opportunities for girls to take on leadership roles in STEM clubs, competitions, and academic projects. Programs that combine academic rigor with mentorship and social belonging have been shown to enhance persistence in advanced STEM coursework (Herrmann et al., 2016). High school interventions should also include opportunities for experiential learning, such as internships, research projects, and industry partnerships to make STEM engagement more concrete and career-oriented (Tai et al., 2006).

Additionally, future research and practice should give greater attention to cultural factors and institutional barriers that shape students' academic experiences. While individual-level interventions such as growth mindset training or exposure to role models have shown promise, they may be insufficient if implemented without consideration of the broader social and structural context in which students learn. As highlighted in the literature, institutional inequities including limited access to advanced coursework, underrepresentation of diverse educators, and a lack of culturally relevant pedagogy reinforce patterns of exclusion and stereotype threat (Nasir et al., 2020). Similarly, cultural values shape how identity and belonging are understood, with Western individualistic framings of interventions potentially limiting their applicability in collectivist contexts (Markus & Kitayama, 1991). Therefore, effective recommendations should move beyond a narrow focus on individual resilience and instead promote systemic reforms that create inclusive learning environments, address racialized and gendered inequities, and validate diverse cultural perspectives in STEM education.

Ultimately, the most effective pathway forward is a multi-tiered, developmentally responsive approach. Early elementary interventions should prioritize concrete, hands-on activities and positive identity formation. These should be followed by scaffolded supports in middle and high school that build on prior knowledge and promote sustained interest. Districts and educational researchers must prioritize the collection of longitudinal data to assess the long-term effectiveness of these interventions, as well as their scalability within diverse school systems (Lipsey et al., 2012). Regular program evaluations and equity audits can ensure that interventions are inclusive, culturally relevant, and equitably distributed.

By adopting these evidence-based practices, schools can create inclusive environments that foster competence, curiosity, and confidence among girls in STEM. Through early intervention, sustained support, and intentional representation, educators can dismantle long-standing barriers and create a more equitable future in science, technology, engineering, and mathematics for all learners.

Limitations and Future Directions

While this meta-analysis offers important insights into the efficacy of STEM interventions for girls, several limitations must be acknowledged to contextualize the findings and guide future research. First, the relatively small number of high-quality, randomized studies in this domain constrains the generalizability of the results. Although the pooled effect sizes suggest consistent benefits across knowledge, interest, and identity domains, the substantial heterogeneity ($I^2 > 70\%$) indicates that intervention effectiveness varies widely depending on delivery context, student age, and intervention type. This variability may reflect unmeasured moderators such as cultural background, classroom climate, or teacher beliefs; factors known to influence educational outcomes, particularly in identity-based interventions (Eccles, 2009; Wang & Degol, 2017).

Another important limitation involves the short-term nature of most included studies. Few interventions assessed long-term retention of STEM interest or achievement gains, limiting conclusions about sustained impact. Only three studies in the dataset conducted delayed follow-ups, and even these were relatively brief (ranging from three to six months). Given that identity development is a longitudinal process shaped by cumulative experiences (Tajfel & Turner, 1986; Dasgupta, 2011), the absence of robust longitudinal data impedes our ability to determine whether observed changes in STEM identity persist or translate into concrete outcomes such as course enrollment or career aspirations. Future studies should adopt multi-year designs with repeated measures to better understand the durability and trajectory of intervention impacts.

Moreover, many interventions focused on girls as a homogeneous group, without disaggregating outcomes by race, socioeconomic status, or intersectional identities. Social Identity Theory emphasizes the dynamic interaction between personal and group identity (Tajfel & Turner, 1986), yet many girls from marginalized racial or linguistic communities may experience STEM in qualitatively different ways due to compounding forms of stereotype threat or systemic exclusion (Cheryan et al., 2011; Steele, 1997). Future research must address these intersectional factors to create inclusive interventions that resonate across diverse populations.

There is also a notable gap in research on scalable, teacher-led models. While expert-delivered interventions were slightly more effective, this delivery mode limits widespread implementation in typical classroom settings. Teacher training in identity-affirming pedagogies and growth mindset principles, when combined with structural support and culturally relevant materials, could help bridge this gap (Darling-Hammond et al., 2020). Evaluations of teacher-led interventions should include fidelity measures to ensure consistency and provide clearer insight into which components are most impactful.

Finally, while the present synthesis employs rigorous inclusion criteria and statistical techniques, it is susceptible to publication bias. Funnel plot asymmetry and trim-and-fill procedures suggest that null or small-effect studies may be underreported. To combat this, future work should emphasize pre-registration, open data, and the publication of null findings, particularly given the growing emphasis on evidence-based policy and practice.

In sum, future research should prioritize longitudinal, intersectional, and scalable studies grounded in strong theoretical frameworks like Social Identity Theory. Doing so will not only clarify which intervention components are most effective but also ensure that strategies for promoting girls' STEM engagement are both equitable and enduring.

Conclusion

In conclusion, this study affirms that targeted STEM interventions can produce meaningful, though modest, gains in girls' knowledge, interest, and identity, especially when grounded in cognitive reframing and implemented early in development. By leveraging evidence-based strategies and the insights of Social Identity Theory, educators and policymakers can design interventions that not only challenge harmful stereotypes but also foster enduring engagement. Continued research is essential to refine these approaches, ensure their inclusivity, and sustain their impact over time.

REFERENCES

American Association of University Women (AAUW). (2020). *The STEM gap: Women and girls science, technology, engineering, and math*. Retrieved from https://www.aauw.org

Aronson, J., & McGlone, M. S. (2009). Social identity and stereotype threat. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 153–178). Psychology Press.

Ball, J. H. (2014). *The impact of stereotype threat on high school females' math performance: Moderators and an intervention* (Publication No. 3662472) [Doctoral dissertation, Louisiana Tech University]. ProQuest Dissertations and Theses Global.

Bancroft, A., Bratter, J., & Rowley, K. (2017). Affirmation effects on math scores: The importance of high school track. *Social Science Research, 64*, 319–333. https://doi.org/10.1016/j.ssresearch.2016.10.001

Banaji, M. R., & Greenwald, A. G. (2013). *Blindspot: Hidden biases of good people*. Delacorte Press.

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math, and engineering majors. *Social Psychology of Education, 15*(4), 427–448. https://doi.org/10.1007/s11218-012-9185-3

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*(2), 256–276. https://doi.org/10.1037/0096-3445.136.2.256

Betz, D. E., & Sekaquaptewa D. (2012). My fair physicist? Feminine math and science role models demotivate young girls. *Social Psychological and Personality Science, 3*(6), 738–746. https://doi.org/10.1177/1948550612440735

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.https://doi.org/10.1002/9780470743386Wiley Online Library+3SCIRP+3SCIRP+3

Borman, G. D., & Grigg, J. (2013). *An effort to close achievement gaps at scale through self-affirmation*. Society for Research on Educational Effectiveness. https://eric.ed.gov/?id=ED563077

Boucher, K. L., Fuesting, M. A., Diekman, A. B., & Murphy, M. C. (2017). Can I work with and help others in this field? How communal goals influence interest and participation in STEM fields. *Frontiers in Psychology, 8*, 901. https://doi.org/10.3389/fpsyg.2017.00901

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour, 5*(8), 980–989. https://doi.org/10.1038/s41562-021-01143-3

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin*, 139(3), 655–701. https://doi.org/10.1037/a0029531

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, 44(8), 1187–1218. https://doi.org/10.1002/tea.20237

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences, 108*(8), 3157– 3162. https://doi.org/10.1073/pnas.1014871108

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135*(2), 218–261. https://doi.org/10.1037/a0014412

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143*(1), 1–35. https://doi.org/10.1037/bul0000052

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology, 97*(6), 1045–1060. https://doi.org/10.1037/a0016239

Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues, 67*(3), 469–491. https://doi.org/10.1111/j.1540-4560.2011.01710.x

Cimpian, J. R., Kim, T., & McCoach, D. B. (2020). Understanding and addressing gender disparities in STEM. *Policy Insights from the Behavioral and Brain Sciences, 7*(1), 20–26. 10.1126/science.aba7377

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*(5791), 1307–1310. https://doi.org/10.1126/science.1128317

Covidence. (2025). *Covidence systematic review software*. Veritas Health Innovation. https://www.covidence.org

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology, 106*(6), 1691–1730. https://doi.org/10.1086/321299

Credé, M. (2018). What shall we do about grit? A critical review of what we know and what we don't know. *Educational Researcher*, 47(9), 606–611. https://doi.org/10.3102/0013189X18801322

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underachievement of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*(6), 588–594. https://doi.org/10.1177/0146167298246003

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review, 43*(6), 1241–1299. https://www.jstor.org/stable/1229039

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development, 82*(3), 766–779. https://doi.org/10.1111/j.1467-8624.2010.01529.xPubMed+2srcd.onlinelibrary.wiley.com+2srcd.onlinelibrary.wiley.com+2

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28(12), 1615–1628. https://doi.org/10.1177/014616702237644

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://doi.org/10.54300/122.311Learning Policy Institute

Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to Equity will determine our future*. Teachers College Press.

Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry, 22*(4), 231–246. https://doi.org/10.1080/1047840X.2011.607313

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*(2), 78–89. https://doi.org/10.1080/00461520902832368Taylor & Francis Online

Eccles, J. S. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioral Development, 35*(3), 195–201. https://doi.org/10.1177/0165025411398185

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11*(2), 135–172. https://doi.org/10.1111/j.1471-6402.1987.tb00781.x

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127. https://doi.org/10.1037/a0018053

Etzkowitz, H., Kemelgor, C., & Uzzi, B. (2000). *Athena unbound: The advancement of women in science and technology*. Cambridge University Press.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532–538. https://doi.org/10.1037/a0015808

Fox, M. F. (2001). Women, science, and academia: Graduate education and careers. *Gender & Society, 15*(5), 654–666. https://doi.org/10.1177/089124301015005002

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Gladstone, J. R., Tallberg, M., Jaxon, J., & Cimpian, A. (2024). What makes a role model motivating for young girls? The effects of the role model's growth versus fixed mindsets about ability and interest. *Journal of Experimental Child Psychology, 238*, Article 105775. https://doi.org/10.1016/j.jecp.2023.105775

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*(1), 17–28. https://doi.org/10.1016/j.appdev.2007.10.004

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*(6), 645–662. https://doi.org/10.1016/j.appdev.2003.09.002

Harter, S. (1999). *The construction of the self: A developmental perspective*. Guilford Press.

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M. C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching, 47*(8), 978–1003. https://doi.org/10.1002/tea.20363

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science, 23*(8), 899–906. https://doi.org/10.1177/0956797611435530

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.

Hertlein, K. (2014). Advantages and disadvantages of technology in relationships: Findings from an open-ended survey. The Qualitative Report, 19, 1-11. Retrieved August 12, 2024, from https://nsuworks.nova.edu/tqr/vol19/iss11/2/

Herrmann, S. D., Adelman, R. M., Bodford, J. E., Graudejus, O., Okun, M. A., & Kwan, V. S. Y. (2016). The effects of a female role model on academic performance and persistence of women in STEM courses. *Basic and Applied Social Psychology, 38*(5), 258–268. https://doi.org/10.1080/01973533.2016.1209757ERIC+1Taylor & Francis Online+1

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. American Association of University Women.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*(3), 175–179. https://doi.org/10.1111/j.0956-7976.2005.00799.x

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher,* *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798SAGE Journals

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013–3000). Institute of Education Sciences, U.S. Department of Education.Institute of Education Sciences

Li, Y., & Bates, T. C. (2019). Does growth mindset predict academic outcomes? *Nature Human Behaviour, 3*, 1009–1010. 10.1037/xge0000669

Li, Y., & Bates, T. C. (2019). You can't change your basic ability, but you work at things, and that's how we get hard things done: Testing the role of growth mindset on response to setbacks, educational attainment, and cognitive ability.*Journal of Experimental Psychology: General, 148*(9), 1640–1655. https://doi.org/10.1037/xge0000669

Lockwood, P. (2006). "Someone like me can be successful": Do college students need same-gender role models? *Psychology of Women Quarterly, 30*(1), 36–46. https://doi.org/10.1111/j.1471-6402.2006.00260.xSAGE Journals+1Exeter Primo+1

Long, J. S., & Fox, M. F. (1995). Scientific careers: Universalism and particularism. *Annual Review of Sociology, 21*, 45–71. https://doi.org/10.1146/annurev.so.21.080195.000401

Master, A., Cheryan, S., & Meltzoff, A. N. (2014). Reducing adolescent girls' concerns about STEM stereotypes: When do female teachers matter? *Revue Internationale de Psychologie Sociale–International Review of Social Psychology*, *27*(3–4), 79–102. https://shs.cairn.info/journal-revue-internationale-de-psychologie-sociale-2014-3-page-79?lang=en

Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology, 108*(3), 424–437. https://doi.org/10.1037/edu0000061

Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology, 160*, 92–106. https://doi.org/10.1016/j.jecp.2017.03.013

Marfo, P., & Okyere, G. A. (2019). The accuracy of effect-size estimates under normal and contaminated normals in meta-analysis. Heliyon, 5, Article e01838. https://doi.org/10.1016/j.heliyon.2019.e01838

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin, 28*(9), 1183–1193. https://doi.org/10.1177/01461672022812004

McKown, C., & Weinstein, R. S. (2008). The development and consequences of stereotype consciousness in middle childhood. *Child Development, 79*(3), 248–266. https://www.researchgate.net/publication/10794937_The_Development_and_Consequences_of_Stereotype_Consciousness_in_Middle_Childhood

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science, 330*(6008), 1234–1237. https://doi.org/10.1126/science.1195996

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18*(10), 879–885. https://doi.org/10.1111/j.1467-9280.2007.01995.x

National Research Council. (2011). *Successful K-12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics.* National Academies Press.

National Science Foundation (NSF). (2021). *Women, minorities, and persons with disabilities in science and engineering: 2021*. Retrieved from https://www.nsf.gov/statistics/women/

National Science Foundation. (2021). *Women, minorities, and persons with disabilities in science and engineering: 2021*. National Center for Science and Engineering Statistics.

Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? *A meta-analysis of experimental evidence. Journal of Applied Psychology, 93*(6), 1314–1334. https://doi.org/10.1037/a0012702

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., ... & Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences, 106*(26), 10593–10597. https://doi.org/10.1073/pnas.0809921106

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83*(1), 44–59. https://doi.org/10.1037/0022-3514.83.1.44

OECD (2014), *PISA 2012 Technical Report*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/6341a959-en.

Omi, M., & Winant, H. (1994). *Racial formation in the United States: From the 1960s to the 1990s* (2nd ed.). Routledge.

Ong, M., Wright, C., Espinosa, L., & Orfield, G. (2011). Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in STEM. *Harvard Educational Review, 81*(2), 172–209. Retrieved from https://otl.wayne.edu/wider/inside_the_double_bind.pdf

Osborne, J. W. (1995). Academics, self-esteem, and race: A look at the underlying assumptions of the disidentification hypothesis. *Personality and Social Psychology Bulletin, 21*(5), 449–455. https://doi.org/10.1177/0146167295215003

Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science*, 26(6), 784–793. https://doi.org/10.1177/0956797615571017

Palid, O., Cashdollar, S., Deangelo, S., Chu, C., & Bates, M. (2023). Inclusion in practice: A systematic review of diversity-focused STEM programming in the United States. *International Journal of STEM Education, 10*(1), Article 2. https://doi.org/10.1186/s40594-022-00387-3

Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLOS ONE, 11*(1), e0146487. https://doi.org/10.1371/journal.pone.0146487

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press

Rattan, A., Good, C., & Dweck, C. S. (2012). “It's ok-not everyone can be good at math”: Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology, 48*(3), 731–737. https://doi.org/10.1016/j.jesp.2011.12.012

Reynolds AJ, Temple JA, Ou SR, Robertson DL, Mersky JP, Topitzes JW, Niles MD. Effects of a school-based, early childhood intervention on adult health and well-being: a 19-year follow-up of low-income families. Arch Pediatr Adolesc Med. 2007 Aug;161(8):730-9. doi: 10.1001/archpedi.161.8.730. PMID: 17679653.

Reynolds, A. J., Temple, J. A., Ou, S.-R., Arteaga, I. A., & White, B. A. B. (2011). *School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups*. Science, 333(6040), 360–364. https://doi.org/10.1126/science.1203618

Riegle-Crumb, C., Moore, C., & Ramos-Wada, A. (2011). Who wants to have a career in science or math? Exploring adolescents' future aspirations by gender and race/ethnicity. *Science Education, 95*(3), 458–476. https://doi.org/10.1002/sce.20431

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440–452. https://doi.org/10.1037/0022-3514.85.3.440

Schmader, T., & Beilock, S. (2012). *An integration of processes that underlie stereotype threat*. In S. T. Fiske & C. N. Macrae (Eds.), The SAGE handbook of social cognition (pp. 34–50). Sage. http://dx.doi.org/10.1093/acprof:oso/9780199732449.003.0003

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336–356. https://doi.org/10.1037/0033-295X.115.2.336

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440–452. https://doi.org/10.1037/0022-3514.85.3.440

Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., ... & Cohen, G. L. (2013). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology, 104*(4), 591–618. https://doi.org/10.1037/a0031495

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To what extent and under which circumstances are growth mindsets important to academic achievement? *Psychological Science*, 29(4), 549–571. https://doi.org/10.1177/0956797617739704

Smeding, A., Dumas, F., Loose, F., & Régner, I. (2013). Order of administration of math and verbal tests: An ecological intervention to reduce stereotype threat on girls' math performance. *Journal of Educational Psychology, 105*(3), 850–860. https://doi.org/10.1037/a0032094

Song, S., Jiang, N., & Ji, L. (2022). Growth mindset training and the effect of math-gender stereotype threat on girl students. *Best Evidence in Chinese Education*, *12*(2), 1629–1634. https://eric.ed.gov/?id=EJ1366241

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4–28. https://doi.org/10.1006/jesp.1998.1373

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist,* 52(6), 613–629. https://doi.org/10.1037/0003-066X.52.6.613

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 34, pp. 379–440). Academic Press. https://doi.org/10.1016/S0065-2601(02)80009-0

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

Tam, H.-L., Chan, A. Y.-F., & Lai, O. L.-H. (2020). Gender stereotyping and STEM education: Girls' empowerment through effective ICT training in Hong Kong. *Children and Youth Services Review, 119*, Article 105624. https://doi.org/10.1016/j.childyouth.2020.105624

Tiedemann, J. (2000). Gender-Related Beliefs of Teachers in Elementary School Mathematics. *Educational Studies in Mathematics*, *41*(2), 191–207. http://www.jstor.org/stable/3483189

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science, 312*(5777), 1143–1144. https://doi.org/10.1126/science.1128690Science

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole. https://www.scirp.org/reference/ReferencesPapers?ReferenceID=757561

Thum, Y. M., & Kuhfeld, M. (2020). NWEA 2020 MAP Growth Achievement Status and Growth Norms for Students and Schools. NWEA Research Report. Portland, OR: NWEA

The Guardian. (2025, February 5). Billionaire's $100 million donation to University of Sydney to boost female STEM participation. *The Guardian*. https://www.theguardian.com/australia-news/2025/feb/05/billionaire-robin-khuda-100m-donation-university-of-sydney-diversity-stem-sector

UNESCO. (2021). *Cracking the code: Girls' and women's education in STEM*. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000260079

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology,* 92(1), 82–96. https://doi.org/10.1037/0022-3514.92.1.82

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology,* 39(5), 456–467. https://doi.org/10.1016/S0022-1031(03)00019-2

Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*(1), 119–140. https://doi.org/10.1007/s10648-015-9355-xPubMed+1Springer Link+1

Welsh JA, Bierman KL, Nix RL, Heinrichs BN. Sustained effects of a school readiness intervention: 5th grade outcomes of the Head Start REDI program. Early Child Res Q. (2020). 4th Quarter;53:151-160. doi: 10.1016/j.ecresq.2020.03.009

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies? *American Psychologist, 75*(9), 1269–1284. https://doi.org/10.1037/amp0000794

Yeager, D. S., Hanselman, P., Walton, G. M., et al. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y

Zhao, F., Zhang, Y., Alterman, V. *et al.* Can Math-Gender Stereotypes Be Reduced? A Theory-

Based Intervention Program with Adolescent Girls. *Curr Psychol* **37**, 612–624 (2018).

https://doi.org/10.1007/s12144-016-9543-y

LIST OF TABLES

Table 1.1 Description of Included Studies.

Author (Year)	Grade Level	Intervention Type	Effect Size *(g or partial η²)*	Setting	Group Size	Administrator
Bancroft (2017)	9	Self-affirmation	$g = 0.42$	Classroom	Whole group	Expert-directed
Master (2014)	10 - 12	Female teachers	$g = 0.15$	Classroom	Whole group	Teacher-directed
Master (2016)	10 - 12	Non-stereotypical classroom	$g = 0.61$	Classroom	Whole group	Teacher-directed
Ball (2014)	9	Gender-fair intervention	$g = 0.05$	Classroom	Whole group	Teacher-directed
Smeding (2013)	8	Verbal–math order	$g \approx 0.24$ (composite effect)	Classroom	Whole group	Teacher-directed
Song (2022)	11	Growth-mindset training	$g \approx 1.00$	Classroom	Whole group	Expert-directed
Zhao (2018)	7 - 8	Self-affirmation; external-attribution training; implicit-attitude change; long-term justice goals; intelligence growth emphasis; role models; comparative thinking	$g = 0.43$(knowledge); $g = -0.92$, $\eta^2 = 0.05$ (Perception)	Classroom	Whole group	Expert-directed
Borman (2013)	7	Self-affirmation	$g = 0.20$ (Interest); 0.094 (Others)	Classroom	Whole group	Expert-directed
Tam (2020)	7 - 8	Online video-making workshop	$g = 0.14$	Online	Whole group	Expert-directed

Author (Year)	Grade Level	Intervention Type	Effect Size *(g or partial η²)*	Setting	Group Size	Administrator
Grant (2023)	5–8	Girls Go Digital (all-girls technology summer camp)	*g* = 0.57 (Interest); 0.02 (Others)	Camp site	Whole group	Expert-directed
Gladstone (2024)	1–3	Role-model exposure	*g* = 0.06 (Identity); 0.35 (Others)	Classroom	Whole group	Expert-directed
Master (2017)	1	Brief learning experience with technology	*g* = 0.83	Classroom	Whole group	Expert-directed

Note. Only the first author's name is listed for each study. All studies employed whole-group delivery. "Expert-directed" interventions were delivered by specialists (e.g., experimenter or STEM educators), whereas "Teacher-directed" interventions were implemented by the regular classroom teacher.

Table 1.2 Outcome Measures and Effect Sizes of Included Studies

Author & Year	Outcome(s)	Measures	Effect Size (*g or partial* η^2)	Confidence Interval	SMD	Origin of Study
Bancroft (2017)	Knowledge of STEM	PSAT Math Exam; STAAR Algebra I Exam	$g = 0.42$	0.36-0.48	0.42	US
Master (2014)	Knowledge of STEM	Custom Computer-Based Assessment	$g = 0.15$	0.27-0.57	0.42	US
Master (2016)	Interest in STEM	STEM Course Enrollment (pre/post intervention comparison)	$g = 0.61$	0.38-0.84	0.61	US
Ball (2014)	Knowledge of STEM	Basic Achievement Skills Inventory – Math Skills Survey	$g = 0.05$	0.22-0.62	0.42	US
Smeding (2013)	STEM Identity; Self-Esteem; Knowledge; Perception Change	Standardized Math/Verbal Tests; Self-report scales	$g \approx 0.24$ (composite effect)	0.18- 0.30	0.24	France
Song (2022)	Knowledge of STEM	Math Test Scores (pre/post comparison)	$g \approx 1.00$	0.90-1.09	0.997	China
Zhao (2018)	Knowledge of STEM Perception change	Recent School-Based Math Tests; 20-item Math-Gender Stereotype Scale	$g =$ 0.43(knowledge); $g = -0.92$, $\eta^2 = 0.05$ (Perception)	0.20-0.66	0.43	China
Borman (2013)	Interest; Identity; Self-Esteem	WKCE Reading, Math; MAP Spring Assessments	$g = 0.20$ (Interest); 0.094 (Others)	0.04- 0.14	0.09	US

Author & Year	Outcome(s)	Measures	Effect Size *(g or partial η²)*	Confidence Interval	SMD	Origin of Study
Tam (2020)	Interest in STEM	PISA 2012 STEM Interest Subscale	$g = 0.14$	0.04-0.24	0.14	Hong Kong
Grant (2023)	Interest; Identity; Self-Esteem	PSAT Math; STAAR Algebra I; Surveys	$g = 0.57$ (Interest); 0.02 (Others)	0.42- 0.72	0.17	US
Gladstone (2024)	Identity; Self-Esteem; Knowledge; Perception Change	Computer-Based STEM Task; Identity & Self-Esteem Scales	$g = 0.06$ (Identity); 0.35 (Others)	0.41- 0.71	0.56	US
Master (2017)	Knowledge of STEM	Technology Motivation & Learning Outcomes (Experimental Task)	$g = 0.83$	0.60-1.06	0.83	US

Note. Only the first author's name is listed for each study. Where exact statistics were not explicitly reported in the original studies, estimates are based on reported group differences or interpreted from test statistics. Multiple effect sizes are separated by semicolons and correspond to the outcomes and measures in the same row. "Composite effect" refers to averaged or generalized effect size across multiple outcomes when individual values were not available.

Table 1.3 Distribution of Studies by Moderator Variables

Moderator Variable	Categories	Number of Studies	Percentage (%)
Intervention Type	Self-Affirmation	3	25%
	Role-Model Exposure	2	17%
	Growth Mindset Training	2	17%
	Classroom/Curricular Reform	3	25%
	Digital/Technology Workshops	2	17%
Delivery Agent	Expert-Directed	9	75%
	Teacher-Directed	3	25%
Grade Level	Elementary (1–5)	2	17%
	Middle School (6–8)	5	42%
	High School (9–12)	5	42%
Study Origin	U.S.	8	67%
	International	4	33%

Note. Percentages are rounded to the nearest whole number.

Table 1.4 Reported Outcome Variables Across Studies

Author (Year)	STEM Knowledge	STEM Interest	STEM Identity	Self-Esteem	Perception Change
Bancroft (2017)	✓	✓		✓	
Master (2014)		✓	✓		✓
Master (2016)	✓	✓	✓		
Ball (2014)	✓		✓	✓	✓
Smeding (2013)	✓				✓
Song (2022)	✓	✓		✓	✓
Zhao (2018)	✓	✓	✓	✓	✓
Borman (2013)				✓	
Tam (2020)	✓	✓	✓		
Grant (2023)	✓	✓	✓		✓
Gladstone (2024)		✓	✓		✓
Master (2017)	✓	✓			

Note. "✓" indicates that the outcome variable was assessed in the study. Outcome variables that were not assessed statistically but were broadly mentioned are also endorsed.

Table 1.5 Studies Excluded from the Meta-Analysis After Full Text Review and Reasons for Exclusion

Excluded Study	Reason for Exclusion
Barbosa, J. F. S., Challco, G. C., & Bittencourt, I. I. (2024). Does gender-stereotyped gamification increase negative thinking? *Interactive Learning Environments, 32*(9), 5632–5659. https://doi.org/10.1080/10494820.2023.2229145	No intervention
Çetinkaya, E., Herrmann, S. D., & Kisbu-Sakarya, Y. (2020). Adapting the values affirmation intervention to a multi-stereotype threat framework for female students in STEM. *Social Psychology of Education, 23*(6), 1587–1607. https://doi.org/10.1007/s11218-020-09594-8	Wrong population (non-K-12)
Cooper, P. K., & Burns, C. (2021). Effects of stereotype content priming on fourth and fifth grade students' gender-instrument associations and future role choice. *Psychology of Music, 49*(2), 246–256. https://doi.org/10.1177/0305735619850624	Wrong study design
Joy, A., Hartstone-Rose, A., Knox, J., Mathews, C. J., Cerda-Smith, J., & Mulvey, K. L. (2025). STEM ability perceptions, basic needs satisfaction, and intrinsic motivation in adolescents: The role of inclusive perceptions in self-determination. *PLoS ONE, 20*(3). https://doi.org/10.1371/journal.pone.0318266	Wrong study design
Kincaid, S. D. (2015). Factors that promote success in women enrolled in STEM disciplines in rural North Carolina community colleges (Doctoral dissertation, Western Carolina University). *ProQuest Dissertations and Theses, 3700881*, 218 pages.	Wrong study design
Krishnannair, S., & Krishnannair, A. (2024). Empowering tomorrow's scientists: 'Girls in Control' workshop promotes STEM education for young girls. *IFAC-PapersOnLine, 58*(25), 37–42. https://doi.org/10.1016/j.ifacol.2024.10.234	Wrong study design

Mousavi, S. M., Salehi, H., Iwatsuki, T., Velayati, F., & Deshayes, M. (2023). Motor skill learning in Iranian girls: Effects of a relatively long induction of gender stereotypes. *Sex Roles, 89*(3–4), 174–185. https://doi.org/10.1007/s11199-023-01398-2	No intervention
Plante, I., de la Sablonnière, R., Aronson, J. M., & Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology, 38*(3), 225–235. https://doi.org/10.1016/j.cedpsych.2013.03.004	No intervention
Schonning, A., & Perez, S. M. (2024). Impact of an immersive engineering program on children's understanding of and interest in engineering: Addressing gender stereotypes. *Journal of Engineering Education, 113*(4), 1226–1244. https://doi.org/10.1002/jee.20617	Wrong population (non-K-12)
Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring = s. *Journal of Personality and Social Psychology, 104*(2), 277–288. https://doi.org/10.1037/a0030461	Wrong population (non-K-12)
Tarbetsky, A. L., Collie, R. J., & Martin, A. J. (2016). The role of implicit theories of intelligence and ability in predicting achievement for Indigenous (Aboriginal) Australian students. *Contemporary Educational Psychology, 47*, 61–71. https://doi.org/10.1016/j.cedpsych.2016.01.002	Wrong study design
Varghese, N., & Kumar, N. (2020). Femvertising as a media strategy to increase self-esteem of adolescents: An experiment in India. *Children and Youth Services Review, 113*, 104965. https://doi.org/10.1016/j.childyouth.2020.104965	No intervention

Appendix

Appendix A

<u>Code Book</u>

Grade-Level – Study includes the grade-level of its participants.

Lower Elementary – Study participants are in grades K-2.

Upper Elementary – Study participants are in grades 3-5.

Middle School – Study participants are in grades 6-8.

High School – Study participants are in grades 9-12.

Gender – Study includes girls only or boys' and girls' participants.

Intervention Type – Study includes a description of how the intervention was delivered.

Adult-Directed – Intervention is delivered by a professional.

Peer-Mediated – Intervention is delivered through peer.

Computer-Based – Intervention is delivered through a computer.

Intervention Administrator – If adult-directed, study includes what type of professional delivered the intervention.

Classroom Teacher – The intervention was delivered by a general education classroom teacher.

Trained Interventionist – The intervention was delivered by a school interventionist or other professional trained to deliver the intervention.

Experimenter – The intervention was delivered by professional conducting research.

Tutor – The intervention was delivered by a tutor.

Group Size – The study includes how many students received the intervention at once.

Individual – The intervention was delivered one-on-one.

Small Group – The intervention was delivered in a group of no more than five students.

Medium Group – The intervention was delivered in a group of more than 6 students.

Dosage – The study includes the total number of hours that participants received the intervention.