Summer 2021

A Meta-Analysis of Youth Gender Prejudice Interventions

Molly Ackerman-Pulliam
San Jose State University

A META-ANALYSIS OF YOUTH GENDER PREJUDICE INTERVENTIONS

A Thesis

Presented to

The Faculty of the Department of Psychology

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Molly Ackerman-Pulliam

August 2021

The Designated Thesis Committee Approves the Thesis Titled

A META-ANALYSIS OF YOUTH GENDER PREJUDICE INTERVENTIONS

by

Molly Ackerman-Pulliam

APPROVED FOR THE DEPARTMENT OF PSYCHOLOGY

SAN JOSÉ STATE UNIVERSITY

August 2021

Gregory J. Feist, Ph.D.	Department of Psychology
Christine Ma-Kellams, Ph.D.	Department of Psychology
Mildred Alvarez, Ph.D.	Department of Psychology

ABSTRACT

A META-ANALYSIS OF YOUTH GENDER PREJUDICE INTERVENTIONS

by Molly Ackerman-Pulliam

This meta-analysis of youth gender prejudice interventions provides an estimate of the overall effectiveness of empirical efforts to reduce expressions of gender prejudice or promote gender egalitarian attitudes among child and adolescent samples. Studies eligible for inclusion were quasi- or fully-experimental designs with child and adolescent participants ranging from 5-17 years of age that quantitatively measured gender prejudice reduction or increases in an egalitarian view of gender. The final sample of 31 studies were located via database searches using both general and specific keywords and Boolean operators. Effect sizes for measures of gender prejudice were calculated using Cohen's *d* as the estimate of effect. A total of 88 effect sizes were retained for analysis. Overall, the interventions had mean and median *d*-values of 0.16 and 0.21, indicating a small positive effect, whereby gender prejudice was reduced. Tests of heterogeneity revealed significant variation among the observed effect sizes, which necessitated the use of random-effects models to potentially identify moderator variables. Publication status and participant age group were identified as significant moderator variables. An analysis of covariance revealed a significant difference in intervention effectiveness between the youngest and oldest participant age groups when the effect of publication status was removed. The results of this meta-analysis will aid researchers in identifying effective and ineffective intervention methods for children and adolescents and may encourage implementation of improved gender prejudice interventions.

ACKNOWLEDGMENTS

The completion of this work was a very long time coming through a multitude of significant lifetime events, so it feels especially poignant to finally write these words.

To my Committee, thank you for your collective knowledge, time, advice, and certainly your patience. Mildred, thank you for being so generous with your kind words and gentle encouragement; that you stuck with me and my work all this time means a great deal to me. Christine, an enormous thank you to you for joining this group without even knowing me and for pushing me to write with greater clarity. Greg, I don't think I could have done this project without you. Your even keel, endless patience, and enduring confidence that this work would get done were guiding beacons for me as I navigated murky and uncharted waters. And Sana, thank you for the selfless giving of your time, insight, and positivity. I truly meant it when I said it felt as if you were my Thesis Fairy Godmother. I wish you nothing but the best as you take on the next chapter of your life.

To my family and friends, thank you for your ceaseless love and support, especially for the last seven years. I always felt that you were behind me 100 %, which I know wasn't always easy to do. Thank you for the emotional support, financial help, and morale boosts you so willingly gave; I am forever grateful to have such a caring support system in my life.

Ember, I love you. You showed such maturity and understanding when our family dynamic changed so I could finish this work. If you remember anything from this time, I hope it is that anything worth doing is worth doing well, and that even if it is something

they are fearful of, a person can still do hard work. Success comes from showing up and trying your best. In part, this work is dedicated to you.

Mom, I love you. Thank you isn't enough, but it's what I've got. Thank you for always loving me exactly the way I am. I have been so lucky to have you as an exemplar of the sort of person I strive to be. You've always shown me how to remain calm and practical when faced with a problem, to be generous without martyrdom, and that being a good worker means showing up to do the work. In part, this work is dedicated to you.

Anthony, I love you, plain and simple. Thank you for the love you share with me, the extra love you showed Ember and me during this project, and the tough love you know I need now and then. Thank you for your full, immediate, and unfaltering support when I told you I wanted to finish school. I could try to list every specific thank you, but the list would be endless. Just know, that the completion of this project—though only a moment in time—was the culmination of a monumental era in my life, and it was made possible by you. I am thankful every day that I have you as a partner, and being your partner makes me want to be the best version of myself. In part, this work is dedicated to you.

Dad, I love you. You are the reason I never stopped thinking about this work. You are the reason I came back to this work. I could let myself be swept away by regrets over not having finished sooner, but I know that you would be proud of me for finishing, no matter how long it took. I made you a promise, and I am so proud of myself for doing what I needed to in order to keep that promise. Dad, my time, my tears, my research, and my degree are dedicated to you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

A Meta-Analysis of Youth Gender Prejudice Interventions

"Is it a boy or a girl?" This seemingly innocuous question, typically asked of anyone announcing the birth of a baby, suggests that a person's gender, defined by the American Psychological Association (n.d.-b) as the implied "psychological, behavioral, social, and cultural aspects of being male or female (i.e., masculinity or femininity)" is a crucial detail. So crucial is this detail, in fact, that Rubin, Provenzano, and Luria (1974) found that within the first 24 hours of life, first-time parents label their newborns differently on the basis of gender, which conceivably foreshadows a lifetime of being perceived and treated in biased ways due to gender alone. Gender serves as a salient and powerful category for individuals and societies. However, the World Health Organization (WHO, n.d.) notes that the social construction of gender produces disparities and discrimination for girls, boys, women, men, and those who identify otherwise as all are respectively subject to cultural, hierarchical, and inequitable notions of appropriateness. The purpose of this thesis is to study child and adolescent gender prejudice interventions via meta-analytic methods in order to determine their overall effectiveness. This research will provide insights about the characteristics of effective interventions and elements of interventions that require revision.

Gender

It is important to further operationalize the term *gender.* Gender is a cultural distinction that refers to a social role comprised of behaviors, thoughts, traits, and expectations about how a given gender should operate in a society, as well as a legal status (*Planned Parenthood*, n.d.). Gender has a hierarchical effect that influences power

and resource distribution differently to men and women, interpersonal interactions, and perceptions about one's self and others (Canadian Institutes of Health Research [CIHR], 2015).

Gender is frequently confounded with sex, which is a biological term that describes the chromosomes, hormones, and genitals. The lack of a consistent, scientific definition of gender may complicate understanding of to what, exactly, gender refers (Palmer & Clegg, 2020). Both gender and sex are generally thought of as binary constructs- girl or boy, woman or man; however, there is wide variety in the expression and experience of these social roles (CIHR, 2015). There are gender-variant people who do not subscribe to either of these groups, as well as intersex people who may possess male and female characteristics such as hormones and genitalia or chromosomes other than the XX and XY pairs.

A person's *gender identity* is the self-identification of being female, male, or some other category (American Psychiatric Association, 2013). A *cisgender* person is someone whose gender identity matches their sex. A *transgender* or *trans* person is someone whose gender identity aligns with that of a different sex from their assigned sex. Additionally, the terms *gender fluid* or *genderqueer* are used to describe gender identities that are neither male nor female but may be some combination of these. Given the spectrum nature of gender, the intention behind this study was to examine gender prejudice interventions regarding attitudes toward girls, boys, women, men, and gender-variant people.

The formation of one's gender identity surely begins with gender assignment, which is the first designation as female or male, usually taking place at birth or before, otherwise known as the "natal gender" (American Psychological Association, n.d.-c.). Researchers have suggested various theories over the course of gender research as to the origins of a person's gender. Following multiple discussions between researchers, Maccoby (1966) compiled chapters on varying gender development theories including hormonal influence on behavior, cognitive developmental theory, and social learning theory. This seminal book spurred the discourse on gender development theories. Current views of gender development maintain that gender develops through an interplay of biological, social, and cognitive mechanisms, though different emphasis is suggested by different researchers as to which facet wields the most influence.

Biological Theory

Wooley (1910) reviewed emerging trends in psychological research and noted that the psychological differences between the sexes were likely due to non-biological factors given the methodological flaws (e.g., small sample sizes, researcher bias) of explanations for differences between males and females. Despite this assessment of biology's minimal impact on mental traits, biological theories continued to underpin psychological research on gender/sex roles through the 1950s and 1960s.

Modern studies of the biological contribution to gender demonstrate the influences of genes and hormones (see Berenbaum et al., 2011 and Polderman et al., 2018 for reviews). One of the basic findings is that there is a heritability component to gender identity. Roughly one-third of a person's gender identity is due to genetic factors, as evidenced by

the systematic review of twin studies by Polderman and colleagues (2018). Another basic finding is that sex hormones have both organizational and activational effects on the brain, meaning that permanent and temporary changes to brain structures and their corresponding behaviors are related to sex hormones. For example, the timing of pubertal onset, as well as the hormonal increases during early-mid puberty have been linked via functional magnetic resonance imaging with risk and reward processing, emotional-cognitive interactions, and emotional and social processing (Herting & Sowell, 2017). However, researchers are still investigating the mechanisms underlying these changes and whether the changes are directly or indirectly influenced by sex hormones (Berenbaum et al., 2011; Herting & Sowell, 2017). Additionally, the mediating and moderating effects of social contexts must be considered.

A wealth of evidence of the biological impact on gender comes from studies of females born with congenital adrenal hyperplasia (CAH), a type of genetic disorder that may result in ambiguous female genitalia and overproduction of androgens, which highlight the effects of hormones on gender-typed behaviors. Females with CAH participate in more male-typed activities over the lifespan, have more male-typed professional interests, and these trends are not related to parental influence or CAH characteristics (Berenbaum et al., 2011). Berenbaum (2018), for example, found that girls with classic CAH (prenatal androgen exposure), as compared to girls with non-classic CAH (postnatal androgen exposure), engaged in significantly more male-typed activities, but the two groups did not differ in their gender identities or cognitions about gender (e.g., gender typicality, or the extent to which a person perceives one's self as a typical

member of the same-sex peer group and gender contentedness, or the extent to which a person is satisfied with their gender assignment). Furthermore, CAH type was not significantly related to time spent with other-gender peers; however, CAH type indirectly affected time spent with same-gender peers via gendered activity interests. That is to say, that girls with classic CAH reported greater interest in male-typed activities, and as a result, they spent less time with same-gender peers due to differing preferences in how to spend time. These findings underline the influence of prenatal androgen exposure on shaping the gender development.

Social Learning Theory

Mischel (1966) conceptualized gender development under a social learning framework that emphasized the role of environmental factors in directing children's behaviors, which leads them to thoughts about appropriate behaviors and the implications of those behaviors (as cited in Martin et al., 2002). A child's appropriate sex-typed behaviors are reinforced by adults, namely parents and teachers, which informs the child's cognitions about their gender identity. For example, Leeb and Rejskind (2004) demonstrated the socializing influence of adults on gendered development in a mutual gaze study with newborns. Both male and female infants prefer and are more likely to look at faces; however, eye contact and mutual gaze are considered to be socially acquired components of the female gender role. Female infants were found to hold a mutual gaze with an adult longer than males at 13-18 weeks old, after no discernable sex difference in gaze when first tested at 13-112 hours old. These findings, combined with evidence regarding the differential treatment of infants on the basis of sex (e.g. Rubin et

al., 1974), suggest that the female infants may have been socialized to meet and maintain eye contact more so than the males.

Further support for the socialization of gender can be found in a longitudinal examination of the effects of center-based care on the gender development of children from 2-5 years old (Bennet et al., 2020). A center-based care environment is one that may make gender especially salient for young children through exposure to greater numbers of same- and other-sex peers, gendered toys, and adults who may intentionally or inadvertently reinforce gender-typing either implicitly or explicitly via many avenues such as word choice, differential treatment, and classroom organization. The researchers hypothesized that recent or current experience with center-based care, as opposed to zero experience, would correlate with higher and earlier levels of gender-typing, as well as higher levels of gender-related knowledge. Children who were enrolled in center-based care at ages 2 and 3 had significantly more same-sex friendships and engaged in more gender-typed play as reported by their mothers than did their peers who enrolled later or never enrolled.

Cognitive Developmental Theory

Kohlberg's (1966) cognitive developmental theory of gender development- based on Piaget's stage-based theory of cognitive development- maintained that a child's growing awareness of their gender identity leads to imitating same-sex models and adopting sex-typed behaviors (as cited in Leaper, 2011). According to this view, beginning around 2-2.5 years old, children use gendered nouns and pronouns (e.g., "boy" and "her") to label people according to gender categories. By 3 years old, children typically assert their

gender identities, predicted by their use of gender labels. From 3-6 years old, children's notions of gender are based primarily on observable characteristics, with stereotypical traits serving as key gender indicators (e.g., long hair for girls, short hair for boys). Following the assertion gender identities, children come to understand that gender is typically stable over time (*gender stability*) and condition (*gender consistency*). These notions culminate as *gender constancy*, or the understanding that gender is an enduring, personal characteristic, which occurs around 6 years old. Slaby and Frey (1975) described this multi-phase progression toward achieving full understanding of gender constancy. Ruble et al. (2007) found that in addition to this pattern of gender development, the relationships between age, stereotype knowledge, the rigidity of gender beliefs, and centrality and evaluation are mediated by increased understanding of gender stability and gender consistency.

Gender Schema Theory

Bem (1983) proposed another cognitive theory of sex-typing whereby children are the driving forces behind their gender development. Using a cognitive structure to encode and organize knowledge, known as a schema, children in a given culture learn what aspects of gender to observe and imitate. Bem noted that the information encoded into children's gender schemas stems from societal practices of male- and femaleness, lending a social learning aspect to gender development. Gender schemas are subject to change, as they create an associative network that guides perception. A child's readiness to sort categorical information initiates the development of a gender schema, and the developing associations between characteristics for a particular gender further guide the child's

perception of schema-relevant information. Sex-typing results from the formation of a gender schema because children make connections between their own gender and what it means to be that gender in their society.

Support for cognition's impact on gender development is found in numerous infant studies. Preferential looking and habituation designs demonstrate that at 3-4 months old, infants can distinguish between female and male faces and that by 6 months old infants make intermodal associations between voices and faces (Quinn et al. 2002; Younger & Fearing, 1999). Ten-month old infants make associations between gendered objects and faces, suggesting early stereotype formation and understanding (Levy & Haaf 1994). These findings imply that infants take an active role in their gender development by seeking and successfully categorizing gender information. Zosuls et al. (2009) further supported the role of cognition in gender development in a study whereby the use of gender labels (e.g., woman, man, lady, guy) between 17 and 21 months old, predicted gender-typed play. This suggests that increased understanding of gender information directs the gendered behavior of children.

Social Cognitive Theory

Bussey and Bandura (1999) outlined gender development as a type of learning that occurs within a social context through the dynamic and reciprocal interplay between a person, their environment, and behaviors. Children's learning of gender-typed behaviors occurs primarily via the modeling of behaviors, with their expectations about the outcomes of performing behaviors informed largely by past experiences, as well as the subjective value placed on the outcomes. The likelihood of continuing or discontinuing a

behavior is reinforced via self-regulatory mechanisms like self-efficacy beliefs and self-sanctioned standards and environmental sources of reinforcement, such as the peer group.

In a study of the impact of media on the pretend play behavior of 3-5 year old girls, Golden and Jacoby (2018) measured Disney princess stereotype perception and awareness and the influence these constructs had on pretend play. Girls modeled specific, feminine body movements (e.g., twirling, dancing, and hand posing) when playing as princesses, but not as other characters. The movements were typically paired with declarations about which princess the child was playing as. The ubiquity of Disney princesses in these girls' lives, combined with the qualities or behaviors the girls ascribed to princesses and the girls' willingness to emulate the characters are taken together to demonstrate social cognitive theory at work.

Impact of Gender

Whether the development of gender identity and conceptions of gender as a social category are synchronistic or causal events has been of great interest to gender development researchers. The relative influences and contributions of biological, social, and cognitive factors to gender development are similarly engaging topics. In any case, one thing can certainly be said about gender—it is an immensely impactful construct across the lifespan and globe. Gender has permeating and lasting effects on perceptions of the self, as well as on perceptions and treatment by others, which may or may not be favorable (Kray et al., 2017; Crocetti et al., 2019; Warren et al., 2019). It is important to note that gender as a cultural classification is an intersectional identity that functions in concert with intersecting characteristics such as race, nationality, social-economic status,

religion, and both customary and legislated practices. The intricacies of these intersecting identities will not be addressed in this paper.

Gender Prejudice

Glick and Hilt (2000) defined *gender prejudice* as, "prejudiced attitudes (*i.e., the attitude that a group deserves lower social status)* based on gender related categorization of people" (p. 197). It is important to further conceptualize what gender prejudice is, how it develops, and how it might be expressed and measured. Defining gender prejudice as an attitude ties in with the larger body of prejudice research. As an attitude, gender prejudice manifests as a negative evaluation of a person or group of people based on their membership or lack thereof to a particular gender group. This attitude is comprised of three components- the affective, which refers to the emotions associated with the person or group; the cognitive, which contains beliefs and stereotypes; and the behavioral, which is typically negative and may include discrimination (American Psychological Association, n.d.-a.). Given these parameters, it seems that gender prejudice and sexism describe the same construct. The two terms are often used interchangeably, as sexism is typically defined as a prejudice or discrimination based on gender-related categorization (Leaper & Brown, 2014). Glick and Hilt used gender prejudice and sexism interchangeably which is commonly seen in research and review literature (2000). However, the tendency to reduce gender and sex to binary categories and the emphasis on biological structures suggested by the root word *sex* compelled the use of gender prejudice in order to more accurately describe the phenomenon under investigation. Similarly, *sexual prejudice* defined by Herek and McLemore (2013) as, "a negative

attitude toward an individual based on her or his membership in a group defined by its members' sexual attractions, behaviors, or orientation" has strong ties to traditional gender role beliefs but is a separate type of prejudice (pg. 311). Despite the interplay of gender expectations and stereotypes and sexual orientation, our examination of gender prejudice interventions will not include interventions relating to sexual minorities, as sexual prejudice is a separate body of research.

Development of Gender Prejudice

As children form cognitions about gender, they also engage in social categorization on the basis of gender. By their third year of life, they can articulate their gender identities, and this leads to identification with a gender group. Tajfel and Turner (1979) demonstrated that the mere assignment of distinct categories can produce prejudice and discrimination. Furthermore, social identity theory maintains that identifying with a social group promotes in-group favoritism in order to sustain self-esteem (Tajfel & Turner, 1986). Members of the out-group are viewed as homogeneous, stereotype-confirming information is more readily attended to and recalled, and stereotype-disconfirming information is ignored, forgotten, or misremembered (Bigler & Pahlke, 2019). Halim et al. (2017) established that children's gender attitudes and cognitions affect intergroup behavior, leading to biased and even discriminatory behaviors towards other-gender peers. Flexible gender cognitions are associated with more favorable ratings of gender out-groups, while rigid gender cognitions are associated with more positive in-group ratings.

Bigler and Liben (2006, 2007) developed developmental intergroup theory (DIT), whereby social categories are used as the basis for children's categorization, leading to stereotypes and prejudice. Gender is made particularly salient as a category due to its perceptual discriminability, the proportional sizes of groups (i.e., the size of a minority group compared to a majority group), the use of explicit gendered language, and children's predisposition to uncover the implicit importance of gendered language. This informs children's categorizations of self and others, which leads to the development of gender stereotypes and prejudices. For example, Hilliard and Liben (2010) tested the effect of gender salience levels on 3-5 year old children's gender stereotyping by manipulating the extent to which preschool teachers accentuated gender in the classroom. They found that children in the high-salience condition had significantly greater gender-stereotyped attitudes and in- and out-group biases, assessed by ratings of same- and other-sex peers and time spent playing with same- and other-sex peers.

Peer Influence

Through early childhood and into adolescence, reinforcements of gendered beliefs are pervasive, and the peer group serves as an enormous source of influence regarding intergroup attitudes. Brechwald and Prinstein (2011) discussed peer influence as it relates to the socialization of beliefs, attitudes, and behaviors, including prejudicial beliefs. Peers serve as emotional and social support for adolescents, but more importantly, they are sources of instruction about how to behave via the dissemination of positive and negative feedback regarding particular behaviors. This feedback influences adolescents' self-concepts and for many, reinforces behaviors and attitudes that reflect social norms. The

implication is that if social norms are traditional with respect to gender, those who adhere to traditional gender expressions will receive positive feedback and those who do not will receive negative feedback, which would most likely be in the form of peer victimization.

Measuring Gender Prejudice

Measuring gender prejudice can be difficult for a variety of reasons. First, the personal natures of the affective and cognitive components of gender prejudice allow for a person to veil their prejudice, especially if it is not deemed socially acceptable to hold or express a prejudicial attitude. Second, a person simply may not be aware of their implicit prejudices despite their activation when the objects of the negative attitudes are presented (Nosek et al., 2007).

There is an abundance of instruments intended to capture evaluations of the social roles held by different gender groups and individuals; some measures assess stereotypes, gender schemas, self-assessments, feminist beliefs, or gender-role attitudes (McHugh & Frieze, 1997). Commonly used measures to assess sexism or gender prejudice include the Attitudes Towards Women Scale for Adolescents (AWSA), Ambivalent Sexism Inventory (ASI), Modern and Old-Fashioned Sexism Scales, Neosexism Scale, and Social Roles Questionnaire (SRQ) (Galambos et al., 1985; Glick & Fiske, 1996; Swim et al., 1995; Tougas et al., 1995; Baber & Tucker, 2006). Bigler (1997) argued that conceptual and methodological flaws, such as clarity about the targets of attitudes (self vs. others), form of sex-typing being assessed (knowledge vs. attitudes), gender domain (job vs. interest), and response options are hindrances to accurate measurement of children's gender-related constructs. In light of these considerations, Liben and Bigler

(2002) developed the Children's Occupations, Activities, and Traits Questionnaire (COAT-AM) to measure attitudes towards others and the sex-typing of the self in occupational interests, activities, and traits. More recent measures that assess gender attitudes among children and adolescents include the Young Children's Reasoning about Gender norms Measure and the Gender-Based Interaction Outcome Expectancies Measure (Conry-Murray & Turriel, 2012; Zosuls et al., 2011).

Conversely, some researchers have developed measures of gender egalitarianism. For example, Beere et al. (1984) constructed the Sex Role Egalitarianism Scale (SRES) to measure "an attitude that causes one to respond to another individual independently of the individual's sex" (p. 564). Though constructed to measure adult attitudes, the SRES has been used with adolescent populations. Other measures of gender egalitarianism used with child populations are the Gender-Equitable Attitude Measure and the Gender Egalitarianism Scale (Rimal et al., 2013; Sadeghi & Agadjanian, 2019).

Global Gender Prejudice

Globally, gender plays a serious role in peoples' lives by affecting rates of poverty, food insecurity, illiteracy, domestic and sexual violence, access to and responsibility for vital resources like water, air pollution deaths, labor force participation and wages, representation in academia, and likelihood of death during an environmental disaster (UN-Women, 2018). Whereas women and girls fare worse than men and boys along all of these dimensions, the World Health Organization notes that men and boys suffer gendered and negative health outcomes related to smoking, taking health and sexual risks, alcohol consumption, and not seeking health care or other help as a result of beliefs

about masculinity (WHO, n.d.). Furthermore, gender variant people who do not align with a society's expectations about gender norms, experience "violence, stigma, and discrimination" (WHO, n.d.).

International bodies such as World Health Organization, the Organisation for Economic Co-Operation and Development, and the United Nations have implemented robust goals, programs, and support in order to achieve gender equality. Despite these efforts, the Social Institutions and Gender Equality Index Global Report (2019) succinctly captures this goliath challenge in noting that "…reforms can have limited traction unless cultural, social and religious norms and structures are taken into account" (para. 3).

Interventions

Within gender prejudice research are experimental efforts to reduce or moderate expressions of gender prejudice, generally with adult samples, but more increasingly with child samples. There is much discussion on what makes for an effective intervention and what hinders such interventions. Bigler and Pahlke (2019) summarized the types of intervention strategies employed to reduce gender prejudice and stereotyping among youth samples. Primarily, researchers utilize counter-stereotypic models to reduce gender stereotype endorsement. Other strategies include explicit anti-bias messages and intergroup contact approaches that focus on the type and quality of interactions between gender groups. However, the researchers point out that as a whole, gender prejudice interventions with children and adolescents are weak in effect, and this is consistent with previous qualitative assessments that characterized gender stereotyping interventions as

having a weak-moderate effect (see Katz, 1986; Serbin & Unger, 1986). The weakness of these interventions may be due in part to the file-drawer problem, gaps in theoretical and conceptual work regarding attitude change, and weak long-term effects (Bigler, 1999). Additionally, Bigler and Pahlke argued that intervention effectiveness is hindered by theoretical misunderstandings or gaps in the origins of gender biases and the causes of children's and adolescents' gender stereotyping and prejudice (2019).

In their review of common gendered issues faced by adolescents, Daniels and Leaper (2011) composed a convincing argument for the need to reduce the prevalence of gender prejudice and stereotyping among adolescents. In the areas of academic achievement, athletics, body-image, sexuality and sexual orientation, friendship intimacy, and aggression and violence, girls and boys face particular issues, but the underlying principle is always the same: behaving in a way that is counter-stereotypical of one's gender is generally met with negativity and opposition from peers. Similarly, Carver et al. (2003) found that gender typicality (perceiving one's self as a typical member of the same-sex peer group), gender contentedness, and felt pressure for gender conformity all contributed to outcomes on measures of psychosocial adjustment. Specifically, children who felt strong pressure and had low gender typicality and contentedness exhibited higher rates of internalizing problems, low self-esteem, and social competence. Similar findings are well documented within the literature (see Egan & Perry, 2001; Yunger et al., 2004). Smith and Leaper (2006), for example, found a positive relationship between self-worth and feelings of gender typicality, which is consistent with findings that children low in gender typicality may report lower self-esteem. More importantly, however, they found that peer

acceptance acts as a mediator between self-worth and gender typicality, meaning that both gender conforming and nonconforming adolescents were the same on measures of self-worth if they felt accepted by their peer group. In their meta-analysis, Schmitt and colleagues (2014) found a significantly negative relationship between perceiving discrimination against one's self or one's ingroup and subsequent measures of psychological well-being, with larger effect sizes existing for children than for adults.

Statement of the Research Question

Gender bias is a deeply rooted adversary to those in pursuit of a world that treats all genders equitably and respectfully. Working to reduce gender bias during early adulthood and beyond is useful but is akin to a reactive approach. The early-life exposure to pervasive and persistent gender stereotypes, prejudice, and discrimination, coupled with the harmful ways in which these phenomena may impact young lives necessitates the proactive development of effective youth gender bias interventions. In light of these considerations, it is imperative that researchers looking to reduce gender bias among children and adolescents know whether, on the whole, gender bias reduction techniques are effective, and if so, to what extent.

This study's aim was to discern the general effectiveness of child and adolescent gender bias interventions. The term *youth* is used to refer collectively to the participants of interest, as studies with child or adolescent samples were the focus of this study. The inclusion of adolescent samples accomplished two goals: it supplemented the number of studies the sampling from child populations (roughly 5-12 years old) and provided an idea of intervention effectiveness of youth from early childhood through late adolescence.

This study was approached from an exploratory standpoint. The aim was to gain an understanding of the current state of youth gender bias reduction research. This meta-analysis evaluated the aggregate effectiveness of quasi- and fully-experimental gender bias reduction interventions. Qualifying criteria for inclusion were that a study utilized youth samples ranging from 5-17 years of age and quantitatively measured a change in gender bias. Unpublished as well as published studies were sought after, as even failures to reject the null contribute useful information to the study of gender bias interventions. This research benefits the larger body of gender bias research by describing the flexibility of gender bias, as it exists among children and adolescents.

Method

Inclusion and Exclusion Criteria

In order to be eligible for inclusion in this study, intervention studies must have had a primary independent variable that consisted of a treatment, program, or strategy intended to reduce biased gender attitudes. This may be a structured program, with multiple components or a multi-day duration, or a less complex treatment, like a single lesson on prejudice or an intergroup contact situation. Given the three components of an attitude—emotions, cognitions, and behaviors—studies that sought to change feelings, stereotypes or beliefs, or behaviors or behavioral intentions as they relate to gender were considered eligible for inclusion. In addition, as a reduction of gender bias may be considered an increase in gender egalitarianism, intervention studies that sought to increase egalitarian gender attitudes were also considered eligible for inclusion.

Eligible studies must also have had a primary dependent variable that assesses a change in participant gender bias. This could have been assessed using within-subjects methods, such as comparisons of pre-treatment scores on a measure of gender prejudice to post-treatment scores, between-subjects methods, such as comparisons of scores on a gender prejudice measure between control and treatment groups, or mixed methods that compared post-treatment scores between control and treatment groups with respect to pre-treatment scores. More concisely, eligible research designs were those that obtained quantitative measures of participants' levels of gender prejudice after some means of trying to reduce them. Neither studies utilizing solely qualitative or observational measures nor correlational studies were considered for inclusion.

Eligible studies must also have sampled from a youth population. Studies with participant ages ranging from 5-17 years were considered for inclusion. It was expected that most, if not all of the studies considered for inclusion used samples from public or private elementary, middle, and high schools.

As I was unable to find any other research that has sought to quantify the overall estimate of the effectiveness of gender prejudice interventions with child and adolescent samples, no temporal limitation on eligible studies was imposed during the search process; on a similar note, no geographical limitations were imposed in order to increase the number of potentially eligible studies.

Search Strategies

The primary search strategy relied on the use of the PsycINFO and Education Resources Information Center (ERIC) article and citation databases. The proposed search strategies for this study originally included additional databases and reverse searching the reference lists of already-identified eligible intervention studies. However, due to time constraints and the large number of studies initially returned from searching PsycINFO and ERIC, the search process was truncated.

Search Terms

Database search terms were identified in a multi-step process. Previously identified intervention studies and relevant literature review articles provided initial keywords. These and related terms were reviewed, respective to each database's thesaurus. Following the advice of Bramer et al. (2018), key concepts of the research question were identified in order to form the following search strategy elements: 1) gender

prejudice/sexism/gender bias/sex bias/gender equality/gender egalitarianism; 2) intervention/experiment/manipulation; 3) youth attitudes; and 4) measurable change. The number of database entries was noted for the thesaurus terms related to these elements, specifically the numbers of entries when terms were denoted as general or subject-specific. For example, the term *sex bias* returned 3,563 entries within the ERIC database when search as a general term and 2,499 entries when searched as a subject term, suggesting that this term could be searched as subject-specific without the concern of missing out on potentially relevant articles. In addition, the term *gender prejudice* returned 73 entries when searched within the PsycINFO database as a general term and 27 entries when searched as a subject-specific term, suggesting that this term ought to be used in a general sense, as further specificity might constrain the search process. As search terms were identified for the four elements described above, they were entered as strings utilizing the AND/OR functions and field codes (e.g., SU- subject, DE- subject exact). Initial database searches were conducted on February 25th and 26th, 2021 and returned 954 and 1,228 results for the PsycINFO and ERIC databases, respectively. Search results were reduced to 366 and 468, respectively, by gleaning subject terms to be omitted via the NOT function. Appendix A shows the complete database search syntaxes used for PsycINFO and ERIC.

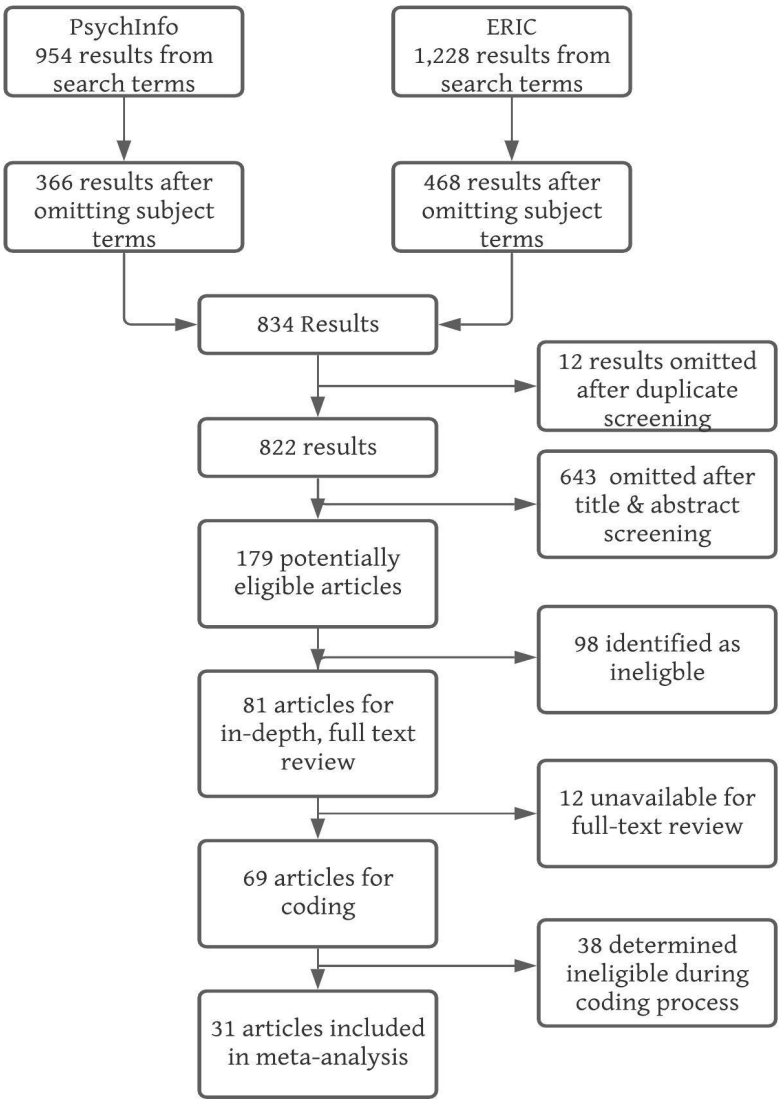
Article Screening and Study Selection

Eight hundred and thirty-four database entries were identified using the methods described above, as illustrated in Figure 1. After screening for duplicate titles, 822 articles were further refined via title and abstract screening, which resulted in 179

potentially eligible articles. These were further reduced to 128 after determining which articles were empirical studies and which reviewed or referenced interventions. Of the remaining articles, 47 were omitted after brief examination of the full text revealed ineligibility for inclusion (e.g., qualitative results, special populations). This left 81 empirical studies, and of these 69 were obtained for coding. During the data extraction process, another 38 studies were omitted from inclusion after reading the full text identified characteristics that precluded inclusion in the meta-analysis (e.g. participants aged younger than 5 or older than 17 or inadequate information with which to calculate effect sizes). This left a final sample of 31 empirical study articles that were included in the meta-analysis, though some articles reported the results of more than one study.

Figure 1

Search Results and Study Selection Process



Grey Literature

The file-drawer problem is a serious potential threat to the strength of a meta-analysis (Rosenthal, 1979). To offset this concern, five primary authors of previously-identified

eligible studies were considered for contacting with inquiries about works in progress, as well as intervention studies that produced nonsignificant results. Presumably current contact information for four authors was available; however, this did not result in the identification of any potentially eligible studies for inclusion in this meta-analysis, as no responses were received.

Coding Procedures

The 31 articles included in this meta-analysis were coded alongside another San Jose State University graduate student. After creating the coding scheme with input from the thesis advisor, the other graduate student was trained on extracting the relevant information from each article. Coders met periodically to discuss key coding decisions and disagreement or confusion regarding coding decisions were resolved via input from the thesis advisor. Two of the articles were jointly coded, nearly one-third of the articles were coded by the other graduate student (n = 8), and the remaining 21 articles were coded by the primary author. In addition to inclusion criteria, additional information was recorded about each reported study, including publication characteristics (e.g. publication year and type), setting and participant details (e.g. geographical region, average socio-economic status (SES) of samples), methodologies (e.g. study design, overall sample size), results, and quality assessment characteristics (e.g. length of intervention, peer-review status). See Tables 1, 2, and 3 for full descriptive statistics of the meta-analysis sample.

Table 1

Publication Characteristics

Characteristic	Coding	*n*	%
Publication Year	1976-1979	5	16.13
	1981-1985	6	19.35
	1990-1999	8	25.81
	2009	3	9.68
	2013-2019	9	29.03
Publication Type	Journal	20	64.52
	Dissertation	10	32.26
	Report	1	3.23
Country	China	2	6.06
	Ethiopia	1	3.03
	Germany	2	6.06
	Germany & Belgium	2	6.06
	Ireland	1	3.03
	Pakistan	1	3.03
	Spain	1	3.03
	Taiwan	1	3.03
	United States	22	66.67
Scope of Bias	Global	18	58.06
	Occupation-based	9	29.03
	STEM-based	4	12.90

Note. n = 31 for each characteristic, except country where n = 33. One study in the sample had participants from more than one country.

Table 2

Participant and Setting Characteristics

Characteristic	Coding	*n*	%
Age (years; months)	5y0m-8y11m	9	27.27
	9y0m-10y10m	7	21.21
	11y0m-13y8m	9	27.27
	14y0m-16y6m	8	24.24
Mean SES	Upper	0	0.00
	Middle	3	9.70
	Low	0	0.00
	Mixed	6	19.35
	Not given	22	70.97
Geographical Area	Urban	3	9.67
	Suburban	6	19.35
	Rural	4	12.90
	Mixed	2	6.45
	Not given	16	51.61
Location	Classroom	27	87.10
	Extracurricular	3	9.67
	Out-of-school	1	3.23

Note. n = 31 for each characteristic, except age where n = 33. Two studies in the sample had participants from multiple age groups.

Table 3

Intervention Characteristics

Characteristic	Coding	*n*	%
Subject Design	Between-subjects	10	32.26
	Within-subjects	2	6.45
	Mixed	19	61.29
Attitude Dimension	Cognitive	24	77.42
	Affective	3	9.67
	Behavioral	1	3.23
	Cognitive & affective	0	0.00
	Cognitive & behavioral	1	3.23
	Affective & behavioral	2	6.45
Target Group of Intervention	Girls/women	12	38.71
	Boys/men	0	0.00
	Girls/women & Boys/men	18	58.06
	Trans	0	0.00
	Any/all	1	3.23
Manipulation Type	Movie	4	12.90
	Curriculum implementation	13	41.94
	Training	5	16.13
	Group discussions	1	3.23
	Role-taking/role-playing	1	3.23
	Informative materials	2	6.45
	Stories	4	12.90
	Other	1	3.23

Note. n = 31 for each characteristic.

Moderator Analyses

It was reasonable to presume that some factors may have moderating effects on gender prejudice intervention effectiveness like participant age, country of intervention, or the experience level of the primary researcher (i.e., whether the research was conducted as part of a degree program). As such, the coding scheme included additional information beyond that required for calculating effect sizes. However, given the exploratory nature of this research, there were no *a priori* predictions about the strength

or directional effects of potential moderator variables. Analyses were conducted with IBM SPSS Statistics (Version 27) and the Meta-Analysis Package for R (Version 3.0-2).

Reducing Prejudice or Increasing Egalitarianism

It was critical to code the directional effect of each outcome included in this study. While all of the interventions were intended to reduce gender prejudice, be it beliefs or stereotypes, feelings about different gender groups, or behaviors related to sexist stimuli, not every intervention functioned to reduce gender prejudice by reducing bias. Some interventions reduced gender prejudice by increasing egalitarianism. This distinction was coded for each outcome as either "reducing prejudice" or "increasing egalitarianism". Aladé (2018), for example, utilized a forced choice outcome measure regarding occupational pictures whereby lower scores for the intervention group on the post-test as compared to pre-test scores would indicate a decrease in traditional or stereotyped attitudes. Thus, this outcome was coded as "reducing prejudice", as the desired outcome was a decrease in intervention group scores. Conversely, Bigler (1991) recorded the egalitarian responses on the Gender Stereotyped Attitude Scale for Children whereby higher post-test scores for the intervention group as compared to the control group indicated a decrease in stereotyping. Thus, this outcome was coded as "increasing egalitarianism", as the desired outcome was an increase in intervention group scores. This coding scheme was necessary to ensure that all effect size calculations were done in such a way that each resulting estimate would accurately reflect whether a given intervention successfully led to a reduction of gender prejudice. In this way, a positive effect size value for a given outcome measure was interpreted as an indication of intervention

success, and a negative effect size value was interpreted as an indication of intervention failure.

Effect Sizes

From the 31 articles, initially 90 different effect sizes were extracted. Cohen's *d* was calculated for each as the estimate of effect size. When a measure of effect size for an outcome measure was directly reported in an article, this value was recorded and used in subsequent calculations and analyses related to that specific outcome. In some instances this required the reported effect size to be transformed to obtain a *d* value. For example, Kwan et al. (2019) reported *d* values as effect sizes which did not require transformation, but Hansen et al. (2014) reported partial eta-squared values as effect sizes which did require transformation. These conversions were done using calculator 14 from Lenton and Lenton's (2016) effect size calculators.

Measures of effect size were not directly reported for the vast majority of the articles. Effect size calculations were computed using various online resources (Lenhard & Lenhard, 2016; Wilson, n.d.; Uanhoro, 2017). When sufficient data were available for pretest-posttest-control group designs, Cohen's *d* values were computed as a difference between the reported pre- and post-test means divided by the pooled standard deviation of both control and treatment groups at pre-test (See Morris, 2008; Lenhard & Lenhard, 2016, Calculator 3). The difference between pre- and post-test means divided by the pooled standard deviation of Time 1 and Time 2 was used for within-subjects designs or control and intervention groups of equal size (Lenhard & Lenhard, 2016, Calculator 1). Calculator 2 from Lenhard and Lenhard calculated *d* in the same way but was used for

post-test only with control designs that had control and treatment groups of different sizes.

When means and/or standard deviations were not reported, effect sizes were determined using recomputation and effect size estimation techniques as outlined in Lipsey and Wilson (2001). For example, *p*-values, as well as *t*, *F*, and *r* statistics can be converted to Cohen's *d*.

Finally, for the articles that reported a nonsignificant effect and did not provide a test result, the effect was coded with a .2 effect size score; this decision was informed by the tendency for reported nonsignificant results to contain a false negative (Hartgerink et al., 2017). Due to the negative skew and non-normal distribution of effect sizes, both the median and mean Cohen's *d* values for all outcomes are reported.

Outlier Analysis

After performing an outlier analysis two effect sizes from Brinkman (2009) were found to be true outliers. After *Z*-score transformation from *d*, the outliers (-4.67 and 7.65, respectively) were more than seven standard deviations away from the mean. This was confirmed visually by plotting the *Z*-scores on a Q-Q plot. As these values were not representative of the literature, they were omitted from further analyses. This left a final sample of 88 effect sizes.

Standard Error and Confidence Intervals

It was necessary to compute the standard error (SE_d) of each effect size estimate. Standard errors were calculated using equations from Hunter and Schmidt (2004) and Becker (1988) (as cited in Nakagawa & Cuthill, 2007). Rosenthal's (1993) conservative

correlation coefficient of 0.7 was used for single group, repeated measures designs, as this figure is commonly not reported. These values were used to calculate the 95% confidence intervals (CI) around each estimate. The CIs were calculated as $d \pm (1.96 \text{ X } SE_d)$.

Results

Descriptive Summary of the Articles

The 31 articles included in the sample were published between 1976 and 2019, with most published between 1990-1999 (k = 8) and 2013-2019 (k = 9). Approximately two-thirds of the sample were peer-reviewed journal articles (k = 20) while the other third were doctoral dissertations (k = 10). The intervention studies spanned 9 countries, though the majority were conducted in the United States (k = 22).

The scope of bias, that is to say the facet of life impacted by gender prejudice that each intervention addressed, was coded as either global, occupation-based, or STEM-based. Interventions with a global scope of bias (k = 18) were those that sought to reduce gender prejudice in an overall sense; occupation-based (k = 9) and STEM-based (k = 4) interventions sought to reduce gender prejudice as it relates to beliefs and stereotypes about professions or future career aspirations.

Girls and/or women and boys and/or men were the target groups of the interventions more often (k = 18) with girls and/or women as the second most common target group (k = 12); interestingly, none of the articles included in the sample specifically sought to reduce gender prejudice toward transgender people. A wide variety of intervention mechanisms were employed, and some of the interventions utilized repeated measures to test the effects of different manipulations. The most common was the implementation of curriculum intended to produce attitudinal change as it relates to gender attitudes (k = 13). See Table 3 for a more detailed summary of the intervention characteristics.

Participant Characteristics

The total number of research participants represented by this meta-analysis was 6,558, with a range of sample sizes from 18 to 1,058 participants. The average sample size was 168 participants per study. The median age of participants was 10 years and 6 months old. As the gender composition of each observed effect size ranged from only boys to only girls, the median gender composition had 53.4% girl participants, giving girls slightly more representation in the overall effect size sample.

Prototype Study

Based on the mode, the typical study was conducted in the United States and published between 1990 and 1999 in a peer-reviewed journal. Gender prejudice was viewed through a global lens as it pertains to cognitions about girls/women and boys/men. The subject design was likely mixed with pre- and post-test measures between a control group and an intervention group. The sample which may have been a nearly even split between girls and boys, if not solely girls, likely had a mean age of nearly 11 years old. The typical intervention took place in a classroom setting and implemented a curriculum to produce attitude change.

Overall Gender Prejudice Intervention Effect Sizes

The mean and median Cohen's *d* values were calculated as 0.16 and 0.21, respectively. The observed range was -1.33 to 2.33 and as illustrated in Figure 2, they appeared to have a mostly normal distribution. Figures 3, 4, and 5 show forest plots of the observed effect sizes and their respective confidence intervals in chronological order.

Figure 2

Histogram of Cohen's d Effect Sizes with Distribution Curve

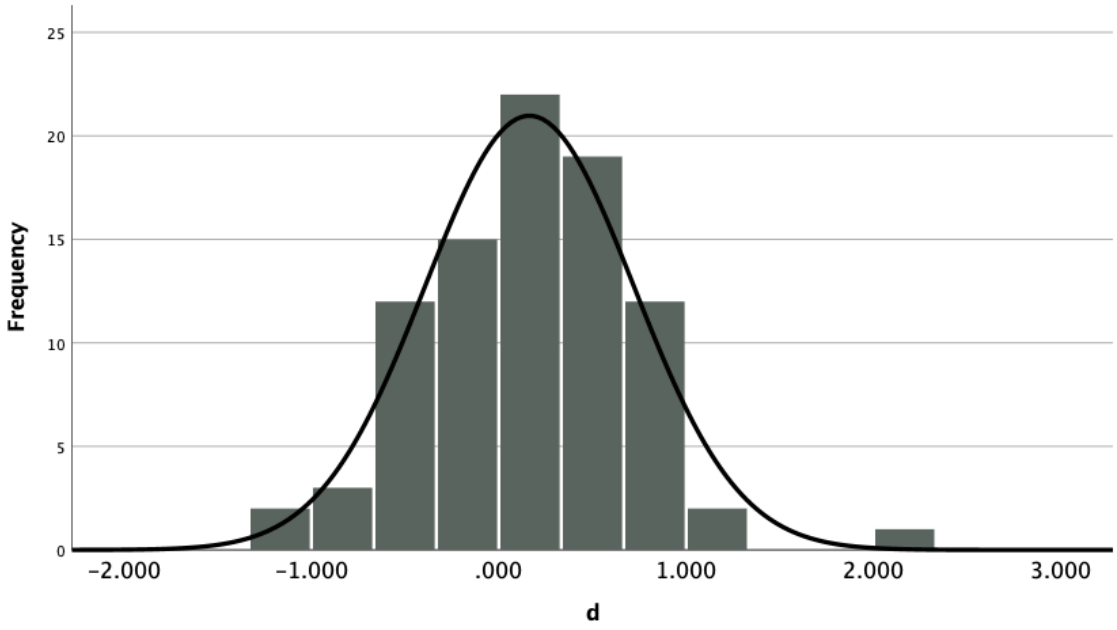


Figure 3

Forest Plot of Confidence Intervals Around Effect Sizes from 1976-1991

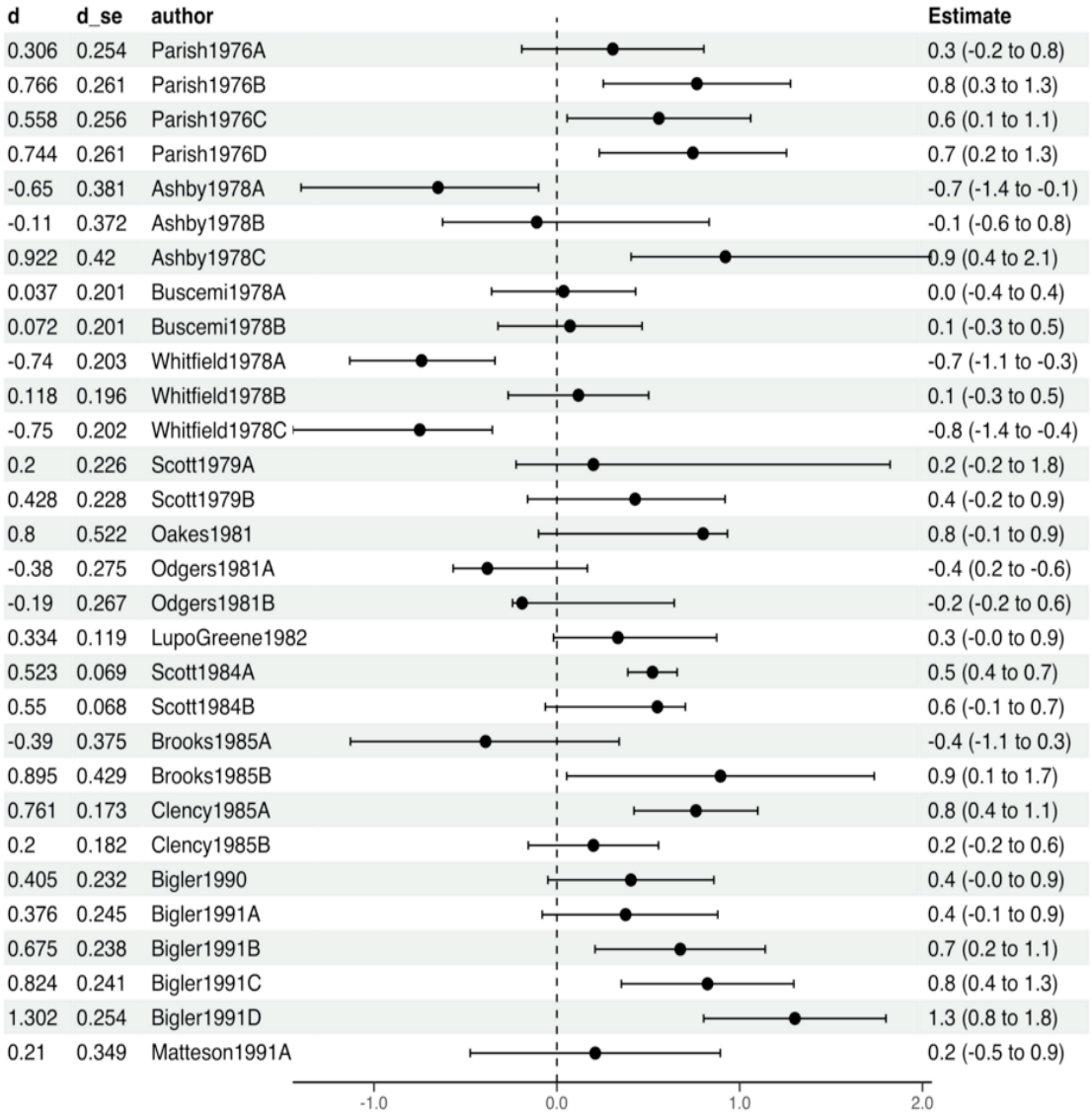


Figure 4

Forest Plot of Confidence Intervals Around Effect Sizes from 1991-2014

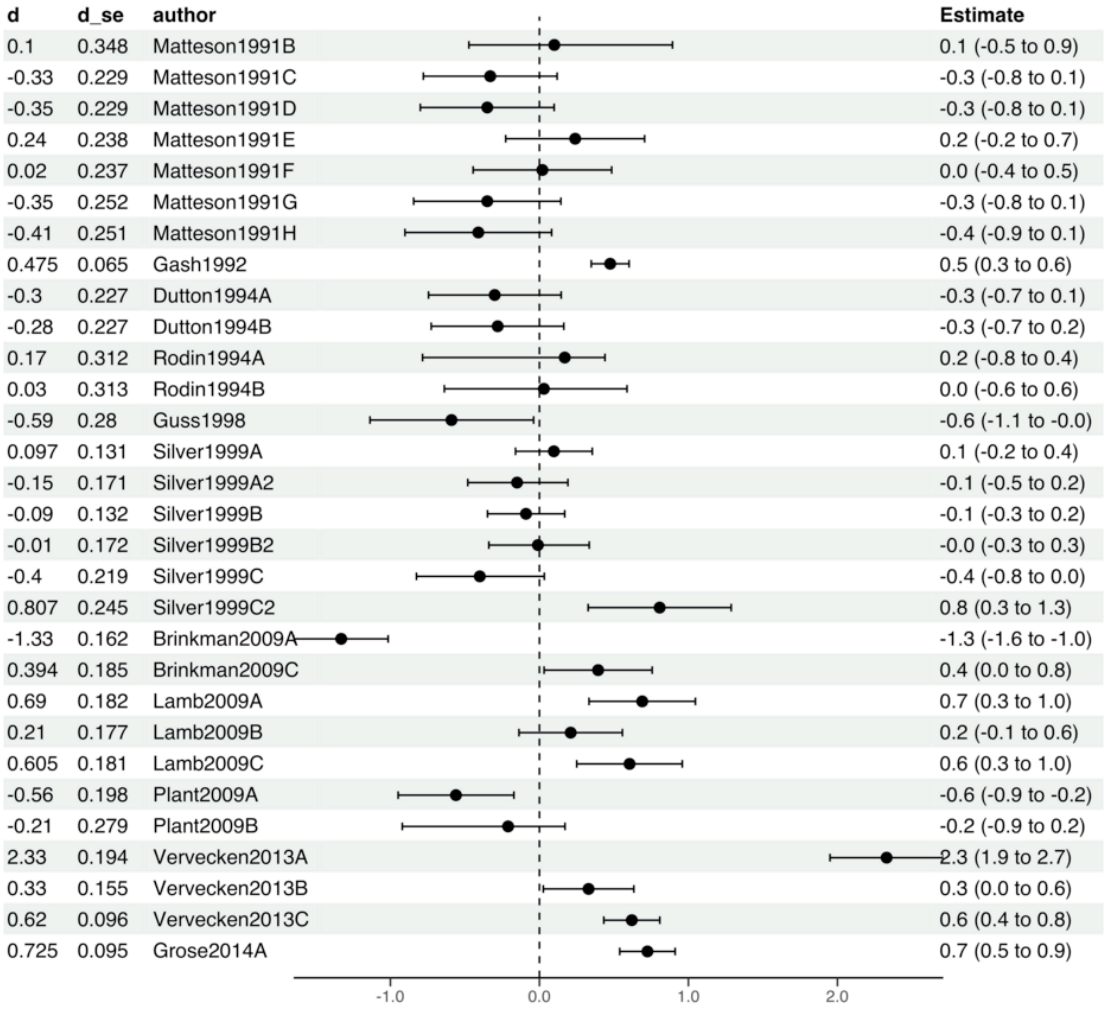
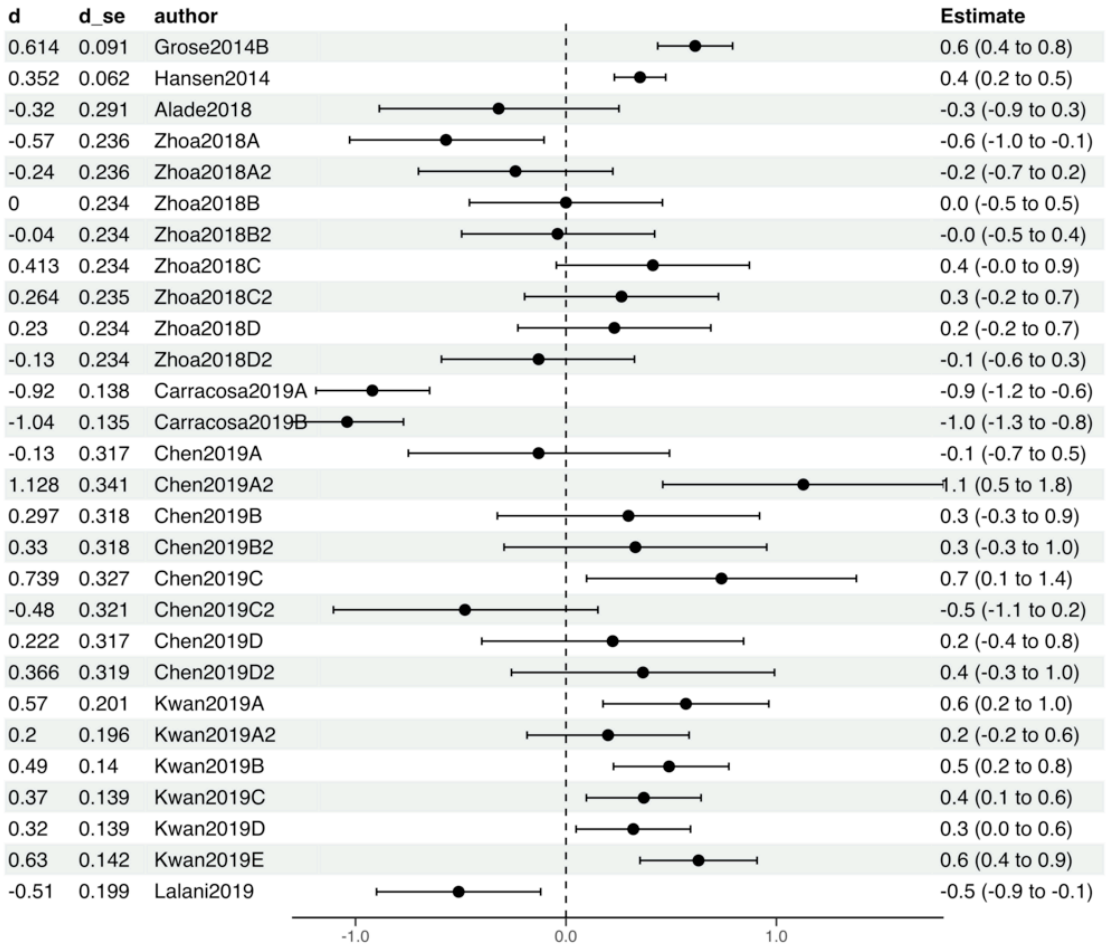


Figure 5

Forest Plot of Confidence Intervals Around Effect Sizes from 2014-2019



Heterogeneity of Effects

Cohen's *d* was used to determine the statistical heterogeneity of the effects. Cochran's *Q* and I^2 are two measures of heterogeneity of effects in meta-analysis (West, et al., 2010, Table 7). Cochran's *Q* is derived from the weighted sum of the squared differences of each study's effect estimate from the pooled effect estimate. If significant—as determined by comparing the statistic with a chi-square distribution with $k - 1$ degrees of freedom (where k is the number of studies)—this test indicates that the

variability of study effects is likely due to something other than mere sampling variability. However, Cochran's *Q* is limited in power when used in a meta-analysis with few studies, and may not detect true heterogeneity (Higgins, et al., 2003).

An alternative measure of heterogeneity of effects is the I^2 statistic, which provides a percentage of variation across studies due to heterogeneity of effect rather than mere sampling error. The primary strength of I^2 as compared to *Q*, is its ability to estimate the impact of heterogeneity on a meta-analysis, rather than simply the presence of heterogeneity. Additionally, I^2 is not constrained by a small number of included studies.

Cochran's *Q* revealed heterogeneity among the effect sizes, Q ($df = 87$) = 136.94, p = 0.0005, which necessitated the use of a random effects model (REM) to calculate a pooled-effects estimate. This type of model assumes that reported or calculated effect sizes exist as a distribution of effect size, rather than a common effect size, which is the underlying assumption of fixed effects models (FEM) (Raudenbush, 1994). The REM test for heterogeneity returned an I^2 value of 38.3%, and the estimate of effect size was d = 0.17, p = 0.0052, CI_{95} [0.05, 0.29]. The I^2 interpretation guideline presented by Deeks et al. (2021) suggests that values under 40.0% indicate heterogeneity may not be of concern. However, Higgins et al. (2003) recommended that values under 25.0% suggest low heterogeneity and values under 50.0% suggest moderate heterogeneity. Taken with the substantial p value of the *Q* statistic, the sample of effect sizes was considered to be markedly heterogeneous. Thus, mixed effects model analyses were performed to investigate the suggested presence of moderator variables. The effect sizes were

subgrouped according to characteristics that could plausibly shed light on the variation of effects.

Some anticipated characteristics guiding subgroupings were mean age or age range of participants, methods used to elicit change in the dependent variable, duration of the interventions, and publication year. Moderation analyses were conducted between effect sizes and suspected moderator variables. Publication year, publication status, and age group were selected as moderator variables for mixed effects modeling as the effect size coding scheme facilitated the formation of these subgroups.

Moderation of Effect Sizes

Publication Year

The literature on gender, gender prejudice, and prejudice intervention has evolved over the years, thus date of publication may elucidate understanding of how intervention effectiveness is influenced by trends in gender research. Publication year did not account for a significant amount of the heterogeneity between effect sizes, Q_M (df = 1) = 0.0136, p = 0.9071. In fact, a two-tailed Pearson correlation demonstrated an inconsequential, negative relationship between publication year and effect size, r = -0.017, p = 0.878. Thus, no further analyses were conducted on the influence of publication year. The test for residual heterogeneity after portioning out publication year, Q_E (df = 86) = 136.87, p = 0.0004 which suggested the presence of other moderator variables.

Publication Status

Publication status associated with peer-reviewed research compared to that of student work (e.g., dissertations) may influence the strength and direction of the relationships

between interventions and their effects. Publication status was determined via publication type, with the one report and 10 dissertations comprising the unpublished group and the 20 journal articles comprising the published group. The amount of heterogeneity accounted for by this model was $R^2 = 13.25\%$, and the test of this variable pointed to significant variation between the published and unpublished effect sizes, $Q_M\,(df = 1) = 4.62$, $p = 0.032$. A point-biserial correlation showed a small, though nonsignificant, negative relationship between effect size and publication status, $r = -0.18$, $p = 0.09$. This was further evidenced by simply comparing the mean Cohen's *d* for the two publication status groups (see Table 4). The test for residual heterogeneity after portioning out publication status was significant, $Q_E\,(df = 86) = 128.29$, $p = 0.0021$ An additional $I^2 =$ 35% of heterogeneity among the effect sizes remained.

Table 4

Descriptive Summary of Effect Sizes by Publication Status

Publication Status	*n*	*M* (*SD*)	*Mdn*	Min	Max
Published	61	0.23 (0.55)	0.31	-1.04	2.33
Unpublished	27	0.01 (0.57)	0.03	-1.33	1.30

Participant Age Group

Participant age was considered as another potential moderator of intervention effect given the wide range of developmental strengths and constraints found from 5-17 years old. Though there was more variation among the age groups than would be expected due to sampling error, $Q_M\,(df = 1)$ 5.11, $p = 0.0237$, a small, but significant negative relationship was found between age group and effect size, $r = -0.26$, $p = 0.014$. In further investigating the relationship between age group and effect size, a univariate analysis of

variance (ANOVA) revealed a nonsignificant effect of age groups. There were no significant differences between the four age groups, F (3, 84) = 2.39, p = 0.075. A simple comparison of the mean Cohen's d for the two publication status groups further underscored this finding (see Table 5). The test for residual heterogeneity after portioning out age group identified persistent heterogeneity, Q_E (df = 86) = 130.1191, p = 0.0015.

Table 5

Descriptive Summary of Effect Sizes by Age Group

Age Group	*n*	*M* (*SD*)	*Mdn*	Min	Max
5y0m-8y11m	21	0.32 (0.54)	0.41	-0.75	1.30
9y0m-10y10m	24	0.31 (0.69)	0.35	-1.33	2.33
11y0m-13y8m	21	0.04 (0.38)	0.04	-0.59	0.76
14y0m-16y6m	22	-0.03 (0.51)	-0.05	-1.04	0.90

Note. Age groups were identified from the spread of mean participant ages across the 88 effect sizes. Four clusters of age ranges were identified, with approximately one-quarter of the effect sizes comprising each cluster.

Exploratory Analysis of Influence of Publication Type on Age Group

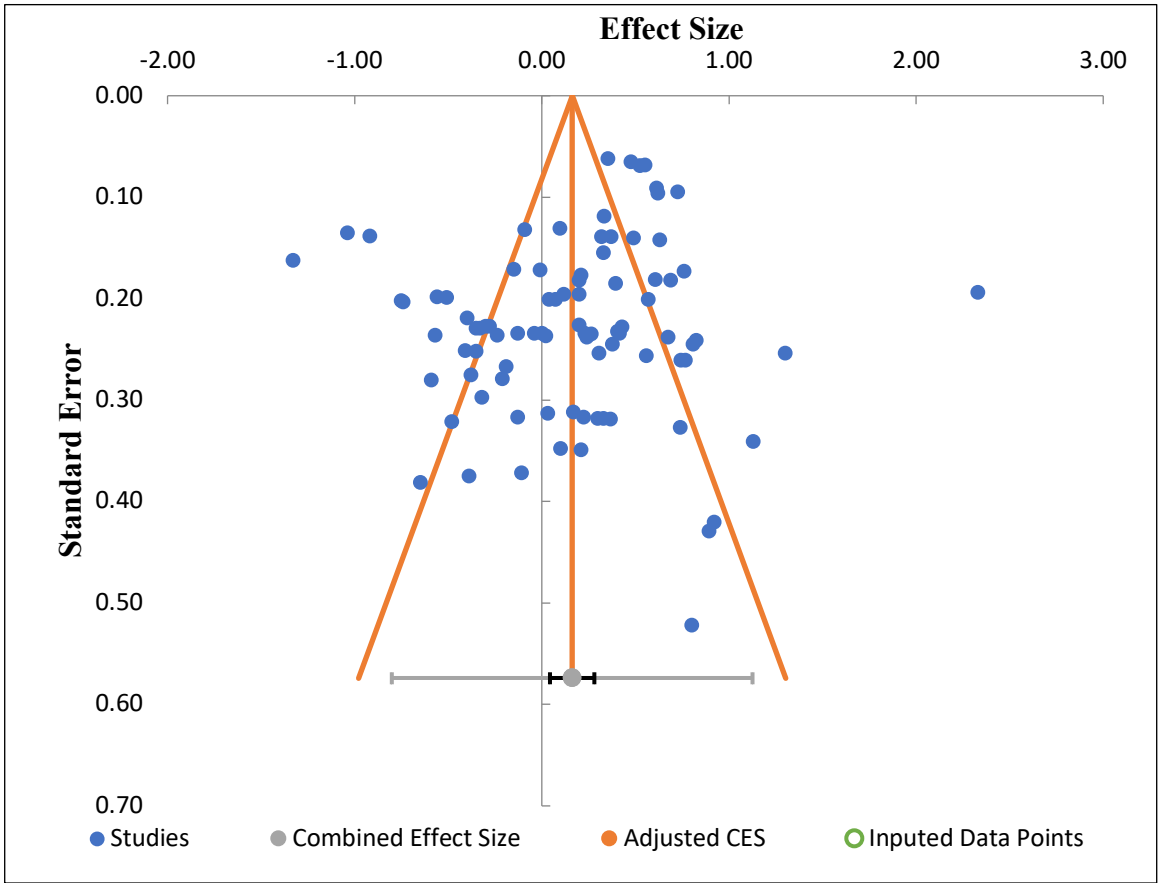
An analysis of covariance (ANCOVA) was conducted to further investigate the difference between age groups while controlling for the influence of publication status. When the influence of publication status was removed, the ANCOVA was significant, F (4, 87) = 2.456, p = 0.05. *Post-hoc* pairwise comparisons revealed a significant difference between the mean effect sizes of the youngest and oldest age groups, p = 0.03. Using a two-tailed Pearson correlation, no correlation between publication status and age group was found, r = -0.01, p = 0.92.

Potential Publication Bias

A potential threat to the validity of this meta-analysis was the risk of a publication bias. A publication bias may have existed as a result of subjectivity by editors and reviewers when evaluating studies on a given topic to publish. Typically, studies with the largest effect sizes are selected for publication, which may lead to misrepresentation and even inflation of the magnitude of effects within the literature; meta-analyses are especially vulnerable to the threat of publication bias as their data sets are largely comprised from published studies. Figure 6 shows the funnel plot generated to determine whether the overall effect was favored by a publication bias of significant results.

Figure 6

Funnel Plot of Effect Sizes



Publication bias is a common problem in meta-analyses, with smaller sample sizes traditionally requiring larger effects to be accepted for publication. In other words, small effect size/non-significant results are likely to end up in the "file drawer." Rosenthal (1979) provided a way to counter the "file-drawer problem" by measuring the number of nonsignificant results that must be unavailable in order to render the overall effect nonsignificant. This "fail-safe number" is obtained by summing the standard normal deviates of each study and dividing by the square root of the number of combined studies. The resulting number reflects how many null results must exist in order to nullify the overall effect. Rosenthal's suggestion for interpreting this number was that the risk of publication bias was low if $N_R > 5k + 10$ (as cited in Fragkos et al., 2014). This technique was used to further inform the presence of a publication bias. The fail-safe number was calculated as $N = 671$, which is larger than the 450 obtained from $5k + 10$, where $k = 31$, indicating minimal risk of publication bias.

A second method to determine publication bias is through regression analysis. If the distribution of effect sizes around their standard error was asymmetrical, a regression analysis was conducted, with sample size entered as the independent variable and unweighted effect size entered as the dependent variable. A regression slope that does not differ significantly from zero is as an indication that sample size does not predict effect size, suggesting there is no existence of a publication bias. Terrin et al. (2005) demonstrated the likelihood of incorrectly identifying asymmetry from a funnel plot, so a modified version of the Egger test as described by Harbord et al. (2006) was used as a complement. Visually, the funnel plot appeared to be asymmetrical, however, the Egger

test for publication bias was not significant, $p = 0.903$ which indicated sufficient symmetry, and as such no further analyses were conducted.

Discussion

Summary of the Findings

The present meta-analysis combined the effects of experimental, gender prejudice interventions among children and adolescents aged 5-17 years old. The interventions had an overall positive effect on the reduction of prejudicial attitudes on the basis of gender, though with a mean effect size $d = 0.16$, the impact of these interventions is small by typical standards. This finding is comparable to other aggregate research on the malleability of prejudicial attitudes. For example, Lenton, Bruder, and Sedikides (2009) found that interventions aimed at lowering the automaticity of reliance on gender stereotypes among adult samples are effective, but only marginally so. Given their qualitative finding coupled with the quantitative findings of this research, it is clear that working to reduce gender prejudice among adults, as well as children and adolecents is more easily said than done. Beelmann and Heinemann (2014) reported an average effect of $d = 0.30$ in their synthesis of standardized intervention programs aimed at promoting positive intergroup attitudes or preventing and reducing intergroup prejudice. This is especially relevant to the current work, as the interventions in their sample utilized participants ranging from 3 ½ years old to 18 years old. Most notable, however, is that gender prejudice interventions were entirely absent from their pool of studies.

The current study addressed gaps in gender prejudice and developmental research by providing an overall estimate of the effectiveness of interventions intended to reduce

gender prejudice among children and adolescents. No other meta-analysis on this question exists and therefore the results of this study will assist researchers and educators in understanding and anticipating the likely outcomes of implementing existing gender prejudice interventions. This information may also inform the development of new approaches to reducing gender prejudice among children and adolescents, as well as encourage the speedy construction and implementation of more effective intervention methods.

Significance of the Moderators

The age group of participants had a significant impact on the spread of observed effect sizes. In general, interventions produced larger effect sizes with younger participants. Gender prejudice interventions produced attitude change in the desired direction from 5 years old until about 10½ -11 years old. The two oldest age groups had miniscule *d* values, and intervention effectiveness trended towards increased gender prejudice for participants 14-17 years old.

This pattern was not entirely unsurprising as the literature indicates developmental trends in child and adolescent attitudinal and behavioral interventions. Beelmann and Heinemann (2014) reported intervention effect sizes for participants aged 3.5 to approximately 8 years old and 8.75-11 years old that were nearly identical to those found in this study (d = 0.28 and 0.35, respectively). Wilson et al. (2003) studied the effectiveness of aggression interventions for children and adolescents and found that after controlling for methodological characteristics, intervention effectiveness did not vary greatly with age. In their systematic review of ethnic prejudice interventions among

children 8 years old or younger, Aboud et al. (2012) found that just under half of the effects were positive; however, they were unable to disaggregate age-related effects further. Interestingly, meta-analytic examinations of the effectiveness of adolescent outgroup attitude interventions show effect sizes like those typically reported for younger children, with *d* values ranging from 0.30-0.52. However, these studies did not include gender attitudes (Beelmann & Heinemann, 2014; Ülger et al., 2019). This lack of gender attitude coverage in the literature only allows for speculation as to why gender prejudice intervention effectiveness seems to be on par with assessments of similar interventions for younger children, but the same pattern is not observed for adolescents. Given the higher rates of effectiveness of adult interventions, the potential for a nonlinear link between age and effect size presents another avenue of investigation.

Publication type—by way of publication status—had a significant contribution to the observed variance in study effect sizes. However, a comparison between the published and unpublished effect sizes did not reveal a significant difference between the two groups. One probable explanation for this finding is the higher representation of published effect sizes in the sample compared to unpublished effect sizes; the published effect sizes outnumbered the unpublished effect sizes two to one. The finding that the average effect sizes from unpublished and published studies were not significantly different from each other was unexpected. This may speak to the rigor with which doctoral candidates conducted their intervention studies, as 10 of the 11 unpublished studies were dissertations. In fact, two of the unpublished studies went on to be published in peer-reviewed journals (Bigler, 1991; Brinkman, 2009). Additionally, the oversight

and influence of dissertation advisors and committee members on the construction and implementation of interventions may have contributed to the caliber of the dissertations. However, it may simply be the case that the vast majority of dissertations are not published as peer-reviewed research, and as such differences between the observed effects of published and unpublished studies eluded detection given the small sample of unpublished works in this meta-analysis.

Even though publication year was not a significant moderator, this finding is illuminating on the state of gender prejudice interventions. As the studies in this meta-analysis represent over 40 years of research, one can only wonder why gender prejudice interventions have not improved their efficacy rates over time. Nearly 30% of the interventions described here were conducted within the last decade, suggesting a resurgence of interest in impacting meaningful social change through addressing gender biases. The advances in rapidly accessing and sharing information made between the mid-1970s and 2019 surely should have led to marked improvements in intervention methodologies. Furthermore, Leaper and Brown (2018) described the growing trend of researching the development of and attitudes toward non-binary and transgender identities. Currently, there are over 50 pieces of legislation at the state level in the United States that relates to gender, from whether transgender children and adolescents should be allowed to play on athletics teams or participate in activities alongside their same-cisgendered peers to the politicization of evidence-based, gender-affirming healthcare (American Civil Liberties Union, 2021). And even though highly cited researchers in the field have been calling for improvements to intervention methods for years, the lack of

modern gender prejudice interventions with improved efficacy rates suggests a larger issue (Bigler, 1995; Leaper & Brown, 2014; Bigler & Pahlke, 2019).

Implications

The results of the meta-analysis contribute to developmental, inter- and intragroup attitude, and gender-focused areas of research. Specifically, this work further informs the theoretical and applied study of reducing gender biases and promoting gender equality among children and adolescents. Though far from conclusive, this study provides a characterization of the gender prejudice interventions that have been conducted over the years. This exploration of intervention effectiveness provides an excellent jumping off point from which deep dives into improved intervention methodologies can begin. Bigler and Pahlke (2019) described the shortcomings of youth-directed gender prejudice interventions, and this study corroborates their arguments about the apparent weaknesses of such interventions. On a more general level, this study may be of interest to educators and policy-makers whose work directly impacts outcomes for children and adolescents. Similarly, systematic reviewers and meta-analysts may find the results of and questions posed by this study intriguing, which could lead to a more thorough investigation of youth gender prejudice interventions.

Limitations

The foremost limitation of this study is that the conclusions to be drawn from it are limited by the methodology. Overall and moderator analyses were conducted on a sample of effect sizes, rather than the customary single effect size per study. A multivariate regression or three-level approach is recommended when aggregating the effect size

estimations of study with multiple outcomes (Cheung, 2019). There are dependencies between multiple effect sizes from a single study that lead to biased estimations of the standard errors, which in turn increases the chance of a Type-I error (Scammacca et al., 2014). Likewise, including single-group designs alongside designs with two or more groups reduces the reliability of the results. The results of this study should be considered in light of these issues.

Another consideration is that the search process for potentially eligible studies was not exhaustive; time and human capital constraints drastically cut short the proposed search process. As such, there is a high likelihood that relevant studies were missed which limited the amount of data available for drawing conclusions about the true state of gender prejudice interventions. Additionally, it is important to note that the findings from this study are inherently constrained by the file drawer problem. Despite attempts to maximize the number of potentially eligible intervention studies by searching prolific databases, including unpublished dissertations, and contacting notable researchers in the field, some number of relevant studies were undoubtedly unknown. There were even some potentially eligible studies that were unavailable. Furthermore, the results of this study only speak to the potential for interventions to successfully lower gender prejudice via the experimental (or quasi-experimental) methods utilized by the studies included in this meta-analysis; correlational reports of decreases in gender stereotyping and prejudice were not addressed by this research.

Variables such as country of origin, theoretical orientation, and study quality are commonly examined as moderating variables in other meta-analytic studies of related

topics, like investigations of child and adolescent intergroup attitude interventions and automatic gender stereotype interventions among adults (e.g., Beelmann & Heinemann, 2014; Lenton et al., 2009). Country of intervention was considered as a possible moderator during data collection as the cultural influence on gender attitudes is undeniable. However, the United States had greater representation in the sample, and this was determined to be a confounding detail that barred country from moderator analyses. As such, this study cannot speak to the cultural or regional influences on intervention effectiveness. Likewise, theoretical and conceptual orientations regarding attitude change underly the specific strategies employed by intervention researchers; thus it is likely these orientations may moderate effect sizes. However, this information was not readily discerned from the articles, and it was thought improper to infer theoretical orientations when they were not explicitly stated.

Last, but not least, this study is limited in its ability to comment on factors underlying the observed effects. A great deal of information was coded for each study in the sample with the intentions to test additional moderator, as well as mediator variables. These analyses would have fleshed out the overall conclusions regarding youth gender prejudice interventions by providing estimates of how various study elements (e.g., scope of bias, manipulation type, target group) influence the malleability of child and adolescent gender attitudes. Unfortunately, the simple fact is that there was not enough time to run every analysis of interest or to adopt a new strategy for combining effect sizes to improve precision of effect size estimates.

Directions for Future Research

Future research should seek to build on the foundation of this study. A replication using the same data, but correcting for methodological issues, would improve the accuracy of statistical analyses. Likewise, this research would benefit from replications with more focused inclusion criteria, such as participant age. As some studies were omitted from inclusion in the meta-analysis due to the presence or possibility of participants either too young or too old, the results of this study were informed by a limited amount of data. Replications may examine the gender prejudice intervention effectiveness for more narrow age groups. This line of research would also benefit from a deeper look into the influences of study characteristics beyond participant age group and publication year and status. Suspected moderator variables of interest may include participant gender, target group of intervention, manipulation type, and attitude dimension, to name a few. Furthermore, the peculiar relationship found between publication status, age group, and effect size should be investigated further.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Aboud, F. E., Tredoux, C., Tropp, L. R., Brown, C. S., Niens, U., Noor, N. M. (2012). Interventions to reduce prejudice and enhance inclusion and respect for ethnic differences in early childhood: A systematic review. *Developmental Review*, *32*(4), 307-336. https://doi.org/10.1016/j.dr.2012.05.001.

*Aladé, F. (2018). *Character portrayals in STEM-focused educational television shows and their impact on children's attitudes towards STEM* (Publication No. 10822173) [Doctoral dissertation, Northwestern University]. Proquest Dissertations & Theses Global.

American Civil Liberties Union (2021, May 7). *Legislation affecting LGBT rights across the country*. ACLU. Retrieved May 16, 2021 from https://www.aclu.org/legislation-affecting-lgbt-rights-across-country

American Psychological Association. (n.d.-a.). Bases of an attitude. In *APA dictionary of psychology*. Retrieved February 21, 2021 from https://dictionary.apa.org/bases-of-an-attitude

American Psychological Association. (n.d.-b.). Gender. In *APA dictionary of psychology*. Retrieved January 28, 2021, from https://dictionary.apa.org/gender

American Psychological Association. (n.d.-c.). Gender assignment. In *APA dictionary of psychology*. Retrieved January 28, 2021, from https://dictionary.apa.org/gender-assignment

American Psychiatric Association (2013). Glossary of technical terms In *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org.libaccess.sjlibrary.org/10.1176/appi.books.9780890425596.GlossaryofTechnicalTerms

*Ashby, M. S. & Wittmaier, B. C. (1978). Attitude changes in children after exposure to stories about women in traditional or nontraditional occupations. *Journal of Educational Psychology*, *70*(6), 945-949.

Baber, K. M., & Tucker, C. J. (2006). The social roles questionnaire: A new approach to measuring attitudes toward gender. *Sex Roles: A Journal of Research*, *54*(7–8), 459–467. https://doi-org.libaccess.sjlibrary.org/10.1007/s11199-006-9018-y

Becker, B. J. (1988). Synthesizing standardized mean change measures. *British Journal of Mathematical and Statistical Psychology*, *41*, 257–278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Beelmann, A., & Heinemann, K. S. (2014). Preventing prejudice and improving intergroup attitudes: A meta-analysis of child and adolescent training programs. *Journal of Applied Developmental Psychology*, *35*(1), 10–24. https://doi.org.libaccess.sjlibrary.org/10.1016/j.appdev.2013.11.002

Beere, C. A., King, D. W., Beere, D. B., & King, L. A. (1984). The Sex-Role Egalitarianism Scale: A measure of attitudes toward equality between the sexes. *Sex Roles: A Journal of Research*, *10*(7–8), 563–576. https://doi.org.libaccess.sjlibrary.org/10.1007/BF00287265

Bem, S. L. (1983). Gender schema theory and its implications for child development: Raising gender-aschematic children in a gender-schematic society. *Signs*, *8*(4), 598-616. http://www.jstor.org/stable/3173685

Bennet, A., Kuchirko, Y., Halim M. L., Costanzo, P. R., & Ruble, D. (2020). The influence of center-based care on young children's gender development. *Journal of Applied Developmental Psychology*, *69*, 101-157. https://doi.org/10.1016/j.appdev.2020.101157

Berenbaum, S. A., Owen Blakemore, J. E., & Beltz, A. M. (2011). A role for biology in gender related behavior. *Sex Roles: A Journal of Research*, *64*, 804-825. https://doi.org.10.1007/s11199-011-9990-8

Berenbaum, S. A., Beltz, A. M., Bryk, K., & McHale, S. (2018). Gendered peer involvement in girls with congenital adrenal hyperplasia: Effects of prenatal androgens, gendered activities, and gender cognitions. *Archives of Sexual Behavior*, *47*, 915-929. https://doi.org/10.1007/s10508-017-1112-4

*Bigler, R. S. (1991). *Cognitive-based training: Reducing gender-stereotyped attitudes and enhancing memory for counterstereotyped stories* [Doctoral Dissertation, Pennsylvania State University]. [ProQuest Information & Learning]. In *Dissertation Abstracts International* (Vol. 52, Issue 4–B, p. 2322).

Bigler, R. S. (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development*, *66*(4), 1072-1087. https://doi.org/10.2307/1131799

Bigler, R. S. (1997). Conceptual and methodological issues in the measurement of children's sex-typing. *Psychology of Women Quarterly*, *21*, 53–69. https://doi.org.libaccess.sjlibrary.org/10.1111/j.1471-6402.1997.tb00100.x

Bigler, R. S. (1999). Psychological interventions designed to counter sexism in children: Empirical limitations and theoretical foundations. In W. B. Swann Jr., J. H. Langlois, & L. A. Gilbert (Eds.), *Sexism and stereotypes in modern society: The gender science of Janet Taylor Spence.* (pp. 129–151). American Psychological Association. https://doi.org.libaccess.sjlibrary.org/10.1037/10277-006

*Bigler, R. S., & Liben, L. S. (1990). The role of attitudes and interventions in gender-schematic processing. *Child Development*, *61*, 1440-1452.

Bigler, R. S., & Liben, L.S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R.V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 39–89). Elsevier Academic Press.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, *16*(3), 162–166. https://doi .org.libaccess.sjlibrary.org/10.1111/j.1467.8721.2007.00496.x

Bigler, R. S., & Pahlke, E. (2019). "I disagree! Sexism is silly to me!" Teaching children to recognize and confront gender biases. In R. K. Mallett & M. J. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors.* (pp. 299–317). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-814715-3.00012-6

Bramer, W. M., de Jonge, G. B., Rethefsen, M. L., Mast, F., & Kleijnen, J. (2018). A systematic approach to searching: an efficient and complete method to develop literature searches. *Journal of the Medical Library Association*, *106*(4), 531-541. https://dx.doi.org/10.5195/jmla.2018.283

Brechwald, W. A., & Prinstein, M. J. (2011). Beyond homophily: A decade of advances in understanding peer influence processes. *Journal of Research on Adolescence, 21,* 161- 179. https://doi-org.libaccess.sjlibrary.org/10.1111/j.15327 .795.2010.00721.x

*Brinkman, B. (2009). *Evaluation of the FAIR program: Teaching gender equality to children*. [Doctoral dissertation, Colorado State University].

*Brooks, L., Holahan, W., & Galligan, M. (1985). The effects of a nontraditional role modeling intervention on sex typing of occupational preferences and career salience in adolescent females. *Journal of Vocational Behavior*, *26*, 264-276. https://doi.org.libaccess.sjlibrary.org/10.1016/0001-8791(85)90005-3

*Buscemi, V. J. (1979). The effectiveness of a curriculum unit designed to modify stereotypic sex role responses of sixth grade students [ProQuest Information & Learning]. In *Dissertation Abstracts International Section A: Humanities and Social Sciences* (Vol. 39, Issue 7–A, p. 4012).

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, *106*(4), 676–713. http://dx.doi.org.libaccess.sjlibrary.org/10.1037/0033-295X.106.4.676

Canadian Institutes of Health Research. (2015, June 17). *Definitions of sex and gender.* Retrieved February 1, 2021 from https://cihr-irsc.gc.ca/e/47830.html

*Carrascosa, L., Cava, M.-J., Buelga, S., & de Jesus, S.-N. (2019). Reduction of sexist attitudes, romantic myths, and aggressive behaviors in adolescents: Efficacy of the DARSI program. *Psicothema*, *31*(2), 121–127. https://doi.org/10.7334/psicothema2018.245

Carver, P. R., Yunger, J. L., & Perry, D. G. (2003). Gender identity and adjustment in middle childhood. *Sex Roles: A Journal of Research*, *49*(3–4), 95–109. https://doi.org.libaccess.sjlibrary.org/10.1023/A:1024423012063

*Chen, M.-H. M. (2019). How biographies of women in science, technology, and medicine influence fifth graders' attitudes toward gender roles. *SAGE Open*. https://doi.org/10.1177/2158244019893704

*Clency, J. S. (1986). *Influence of a sex equity and a science-mathematics career orientation intervention paradigm on middle school/junior high school students* [Doctoral dissertation, Indiana University]. Dissertation Abstracts International Section A: Humanities and Social Sciences (Vol. 46, Issue 9–A, p. 2540).

Conry-Murray, C., & Turiel, E. (2012). Jimmy's baby doll and Jenny's truck: Young children's reasoning about gender norms. *Child Development*, *83*(1), 146–158. https://doi.org.libaccess.sjlibrary.org/10.1111/j.1467-8624.2011.01696.x

Crocetti, E., Moscatelli, s., Kaniušonytė, G., Meeus, W., Žukauskienė, R., & Rubini, M. (2019). Developing morality, competence, and sociability in adolescence: A longitudinal study of gender differences. *Journal of Youth and Adolescence*, *48*(5), 1009-1021. https://doi.org/10.1007/s10964-019-00996-2

Daniels, E. A., & Leaper, C. (2011). Gender issues. In R. J. R. Levesque (Ed.). *Encyclopedia of adolescence* (Vol. 1, pp. 151-159). Springer.

Deeks, J.J., Higgins, J.P.T., Altman, D.G. (2021). Analysing data and undertaking meta-analyses. In J. P. T Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (*version 6.2*). Cochrane. http://www.training.cochrane.org/handbook.

*Dutton, D. M. (1994). *Girls' and boys' beliefs about adult work: Evaluating the effectiveness of a school-based program.* [Doctoral dissertation, University of Rhode Island]. Dissertation Abstracts International: Section B: The Sciences and Engineering (Vol. 56, Issue 1–B, p. 0564).

Egan, S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology*, *37*(4), 451-463. https://doi-org.libaccess.sjlibrary.org/10.1037/0012-1649.37.4.451

Fragkos, K. C., Tsagris, M., & Frangos, C. C. (2014). Publication bias in meta-analysis: Confidence intervals for Rosenthal's fail-safe number. *International Scholarly Research Notices*, *Volume 2014*, Article ID 825383. http://dx.doi.org/10.1155/2014/825383

Galambos, N. L., Petersen, A. C., Richards, M., & Gitelson, I. B. (1985). The Attitudes Toward Women Scale for Adolescents (AWSA): A study of reliability and validity. *Sex Roles: A Journal of Research*, *13*(5–6), 343–356. https://doi.org.libaccess.sjlibrary.org/10.1007/BF00288090

*Gash, H., & Morgan, M. (1993). School-based modifications of children's gender-related beliefs. *Journal of Applied Developmental Psychology*, *14*(2), 277–287. https://doi-org.libaccess.sjlibrary.org/10.1016/0193-3973(93)90037-V

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491–512. https://doi.org.libaccess.sjlibrary.org/10.1037/0022.3514.70.3.491

Glick, P., & Hilt, L. (2000). Combative children to ambivalent adults: The development of gender prejudice. In Eckes, T., & Trautner, H. M. (Eds.) (2000). *The developmental social psychology of gender.* Lawrence Erlbaum Associates.

Golden, J. C., & Jacoby, J. W. (2018). Playing princess: Preschool girls' interpretations of gender stereotypes in Disney Princess media. *Sex Roles: A Journal of Research*, *79*(56), 299-313. http://dx.doi.org.libaccess.sjlibrary.org/10.1007/s11199-017-0773-8

*Greene, A. L., Sullivan, H. J., & Beyard-Tyler, K. (1982). Attitudinal effects of the use of role models in information about sex-typed careers. *Journal of Educational Psychology*, *74*(3), 393-398.

*Grose, R. G., Grabe, Shelly, & Kohfeldt, D. (2014). Sexual education, gender ideology, and youth sexual empowerment. *Journal of Sex Research*, *51*(7), 742-75. https://doi.org/10.1080/00224499.2013.809511

*Guss, T. O., & Adams, L. (1998). Gender orientation and career maturation among rural elementary school students.

Halim, M. L. D., Ruble, D. N., Tamis, L. C. S., Shrout, P. E., & Amodio, D. M. (2017). Gender attitudes in early childhood: Behavioral consequences and cognitive antecedents. *Child Development*, *88*(3), 882–899. https://doi-org.libaccess.sjlibrary.org/10.1111/cdev.12642

*Hansen, N., Postmes, T., Tovote, K. A., & Bos, A. (2014). How modernization instigates social change: Laptop usage as a driver of cultural value change and gender equality in a developing country. *Journal of Cross-Cultural Psychology*, *45*(8), 1229-1248. https://doi.org/10.1177/0022022114537554

Harbord, R. M., Egger, M., & Sterne, J. A. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine*, *25*(20), 3443–3457. https://doi.org/10.1002/sim.2380

Hartgerink, C. H. J., Wicherts, J. M., & van Assen, M. A. L. M. (2017). Too good to be false: Nonsignificant results revisited. *Collabra: Psychology*, *3*(1). https://doi.org/10.1525/collabra.71

Herek, G. M., & McLemore, K. A. (2013). Sexual prejudice. *Annual Review of Psychology*, *64*, 309–333. https://doi-org.libaccess.sjlibrary.org/10.1146/annurevpsych-113011-143826

Herting, M. M., & Sowell, E. R. (2017). Puberty and structural brain development in humans. *Frontiers in Neuroendocrinology*, *44*, 122–137. https://doi.org/10.1016/j.yfrne.2016.12.003

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, *327*(7414), 557–560. https://doi-org.libaccess.sjlibrary.org/10.1136/bmj.327.7414.557

Hilliard, L. J., & Liben, L. S. (2010). Differing levels of gender salience in preschool classrooms: Effects on children's gender attitudes and intergroup bias. *Child Development*, *81*(6), 1787–1798. https://doi.org.libaccess.sjlibrary.org/10.1111/j.1467.8624.2010.01510.x

Hunter, J. E. & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research finding,* 2nd edition. Sage.

Katz, P. A. (1986). Modification of children's gender-stereotyped behavior: General issues and research considerations. *Sex Roles: A Journal of Research*, *14*(11–12), 591–602. https://doi-org.libaccess.sjlibrary.org/10.1007/BF00287690

Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82-173). Stanford University Press.

Kray, L. J., Howland, L., Russell, A. G., & Jackman, L. M. (2017). The effects of implicit gender role theories on gender system justification: Fixed beliefs strengthen masculinity to preserve the status quo. *Journal of Personality and Social Psychology*, *112*(1), 98-115. http://dx.doi.org/10.1037/pspp0000124

*Kwan, K. M W., Shi, S. Y., Nabbijohn, A. N., MacMullin, L. N., VanderLaan, D. P., & Wong, W. I., (2019). Children's appraisals of gender nonconformity: Developmental pattern and intervention. *Child Development*, *91*, e780-e798. https://doi.org/10.1111/cdev.13316

*Lalani, M. & Bhutta, S. M. (2019). Reducing attitudes of prejudice and discrimination through literature: a quasi-experiment on the upper key-stage 2 students of the British schools in Karachi, Pakistan. *Intercultural Education*, *30*(1), 83-100. https://doi.org/10.1080/14675986.2018.1528772

*Lamb, L. M., Bigler, R. S., Liben, L. S., & Green, V. A. (2009). Teaching children to confront peers' sexist remarks: Implications for theories of gender development and educational practice. *Sex Roles*, *61*, 361-382. https://doi.org/10.1007/s11199-009-9634-4

Leaper, C. & Bigler, R. S. (2011). Gender. In M. K. Underwood & L. H. Rosen (Eds.), *Social development: Relationships in infancy, childhood, and adolescence* (pp. 298-315). Guilford Press.

Leaper, C., & Brown, C. S. (2014). Sexism in schools. In L. S. Liben & R. Bigler (Eds.), *Advances in child development and behavior, Vol. 47: The role of gender in educational contexts and outcomes.* Elsevier Academic Press. http://dx.doi.org.libaccess.sjlibrary.org/10.1016/bs.acdb.2014.04.001

Leaper, C., & Brown, C. S. (2018). Sexism in childhood and adolescence: Recent trends and advances in research. *Child Development*, *12*(1), 10-15. https://doi.org/10.1111/cdep.12247

Leeb, R. T., & Rejskind, F. G. (2004). Here's looking at you, kid! A longitudinal study of perceived gender differences in mutual gaze behavior in young infants. *Sex Roles: A*

Journal of Research, *50*(1-2), 1–5. https://doi.org/10.1023/B:SERS.0000011068.42663.ce

Lenhard, W. & Lenhard, A. (2016). *Calculation of effect sizes*. Psychometrica. https://www.psychometrica.de/effect_size.html

Lenton, A. P., Bruder, M., & Sedikides, C. (2009). A meta-analysis on the malleability of automatic gender stereotypes. *Psychology of Women Quarterly*, *33*(2), 183–196. https://doi-org.libaccess.sjlibrary.org/10.1111/j.1471-6402.2009.01488.x

Levy, G. D., & Haaf, R. A. (1994). Detection of gender-related categories by 10-month-old infants. *Infant Behavior & Development*, *17*(4), 457–459. https://doi.org.libaccess.sjlibrary.org/10.1016/0163-6383(94)90037-X

Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, *67*(2), vii-147. https://doi.org.libaccess.sjlibrary.org/10.1111/1540-5834.t01-1-00187

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.

Maccoby, E. E. (Ed.). (1966). *The development of sex differences.* Stanford University Press.

*Matteson, D. R. (1991). Attempting to change sex role attitudes in adolescents: Explorations of reverse effects. *Adolescence*, *26*(104), 885–898.

McHugh, M. C., & Frieze, I. H. (1997). The measurement of gender-role attitudes: A review and commentary. *Psychology of Women Quarterly*, *21*(1), 1–16. https://doi.org.libaccess.sjlibrary.org/10.1111/j.1471-6402.1997.tb00097.x

Mischel, W. (1966) A social-learning view of sex differences in behavior. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 57–81). Stanford University Press.

Morris, S. B. Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364-386. https://doi.org/10.1177/1094428106291059

Nakagawa, S. & Cuthill, I. C. (2007). Effect size, confidence interval, and statistical significance: a practical guide for biologists. *Biological Review*, *82*, 591-605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88. https://doi.org.libaccess.sjlibrary.org/10.1080/10463280701489O5

*Oakes, J. H. & Williamson, P. A. (1981). Short-term treatment effects on sex-role attitudes. *The Journal of Social Psychology*, *115*, 143-144. https://doi.org/10.1080/00224545.1981.9712001

*Odgers, C. R. (1981). *Development of a workshop which confronts sex bias in vocational education and investigation of the impact of workshop participation on tenth grade students*. [Doctoral dissertation, Ohio State University].

OECD. (2019). *SIGI 2019 Global Report: Transforming Challenges into Opportunities*. OCED Publishing. https://doi.org/10.1787/bc56d212-en.

Palmer, B. F., & Clegg, D. J. (2020). A universally accepted definition of gender will positively impact societal understanding, Acceptance, and appropriateness of health care. *Mayo Clinic Proceedings. 95*(10), 2235–2243. https://doi.org/10.1016/j.mayocp.2020.01.031

*Parish, T. S. & Bryant, W. T. (1976). Reversing the effects of sexism in elementary school girls through the use of counterconditioning procedures. *Journal of Classroom Interaction*, *11*(2), 7-15. Retrieved May 10, 2021, from http://www.jstor.org/stable/23869244

Planned Parenthood. (n.d.). *Sex and Gender Identity*. Retrieved January 21, 2021. https://www.plannedparenthood.org/learn/gender-identity/sex-gender-identity

*Plant, E. A., Baylor, A. L., Doerr, C.E., Rosenberg-Kima, R. B. (2009). Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers and Education*, *53*(2), 209-215. https://doi.org/10.1016/j.compedu.2009.01.013

Polderman, T. J. C., Kreukels, B. P. C., Irwig, M. S., Beach, L., Chan, Y., Derks, E. M., Esteval, I., Ehrenfeld, J., Den Heijer, M., Posthuma, D., Raynor, L., Tishelman, A., Davis, L. K. (2018). The biological contributions to gender identity and gender diversity: Bringing data to the table. *Behavior Genetics*, 48, 95-108. https://doi.org/10.1007/s105190189889z

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*(9), 1109-1121. http://dx.doi.org.libaccess.sjlibrary.org/10.1068/p3331

Raudenbush, S. W. Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-321). Russell Sage Foundation.

Rimal, R. N., Figueroa, M. E., & Storey, J. D. (2013). Character recognition as an alternate measure of television exposure among children: Findings from the Alam Simsim program in Egypt. *Journal of Health Communication*, *18*(5), 594–609. https://doi.org.libaccess.sjlibrary.org/10.1080/10810730.2012.743625

*Rodin, S. B. (1994). *Challenging sex-role stereotypes in early adolescent girls: A group intervention approach*. [Doctoral dissertation, California Institute of integral Studies].

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rosenthal, R. (1993). Meta-analytic procedures for social research. Sage Publications. http://nrs.harvard.edu/urn-3:HUL.InstRepos:38784661

Rubin, J. Z., Provenzano, F. J., & Luria, Z. (1974). The eye of the beholder: Parents' views on sex of newborns. *American Journal of Orthopsychiatry*. *44*(4), 512-519. https://doiorg.libaccess.sjlibrary.org/10.1111/j.1939-0025.1974.tb00905.x

Ruble, D. N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., Shrout, P. E. (2007). The role of gender constancy in early gender development. *Child Development*, *78*(4), 1121-1136. https://doi.org/10.1111/j.1467-8624.2007.01056.x

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research*, *84*(3), 328–364. https://doi.org/10.3102/0034654313500826

Sadeghi, R., & Agadjanian, V. (2019). Attitude and propensity to divorce in Iran: Structural and ideational determinants. *Journal of Divorce & Remarriage*, *60*(6), 479–500. https://doi.org.libaccess.sjlibrary.org/10.1080/10502556.2019.1586228

Schmitt, M. T., Branscombe, N. R., Postmes, T., & Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: A meta-analytic review. *Psychological Bulletin*, *140*(4), 921–948. https://doi.org.libaccess.sjlibrary.org/10.1037/a0035754

*Scott, K. P. & Feldman-Summers, S. (1979). Children's reactions to textbook stories in which females are portrayed in traditionally male roles. *Journal of Educational Psychology*, *71*(3), 396-402. https://doi.org/10.1037/0022-0663.71.3.396

*Scott, K. P. (1984). Effects of an intervention on middle school pupils' decision making, achievement, and sex role flexibility. *The Journal of Educational Research*, *77*(6), 369-375. Retrieved May 10, 2021, from http://www.jstor.org/stable/27540079

Serbin, L. A., & Unger, R. K. (1986). Social change: introduction. *Sex Roles: A Journal of Research*, *14*, 561–566. https://doi.org.libaccess.sjlibrary.org/10.1007/BF00287687

*Silver, B. E. (1999). *A media skills intervention for adolescents on gender attitudes, beliefs, and behaviors*. [Doctoral dissertation, University of Rhode Island]. (Paper 1078). Open Access Dissertations. https://doi.org/10.23860/diss-silver-barbara-1999

Slaby, R. G., & Frey, K. S. (1975). Development of gender constancy and selective attention to same-sex models. *Child Development*, *2*, 849-856. https://doi.org/10.2307/1128389

Smith, T. E., & Leaper, C. (2006). Self-perceived gender typicality and the peer context during adolescence. *Journal of Research on Adolescence*, *16*(1), 91–103. https://doi.org.libaccess.sjlibrary.org/10.1111/j.1532-7795.2006.00123.x

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199–214. https://doi-org.libaccess.sjlibrary.org/10.1037/00223514.68.2.199

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–49). Brooks/Cole.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup* relations (pp. 7–24). Nelson Hall.

Terrin, N., Schmid, C. H., & Lau, J. (2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology*, *58*(9), 894–901. https://doi.org/10.1016/j.jclinepi.2005.01.006

Tougas, F., Brown, R., Beaton, A. M., & Joly, S. (1995). Neosexism: Plus ça change, plus c'est pareil. *Personality and Social Psychology Bulletin*, *21*(8), 842–849. https://doi.org.libaccess.sjlibrary.org/10.1177/0146167295218007

Uanhoro, J. O. (2017). Effect size calculators. *Available online at https://effect-size calculator.herokuapp.com/*

Ülger, Z., Dette-Hagenmeyer, D. E., Reichle, R., Gaertner, S. L. (2018). Improving outgroup attitudes in schools: A meta-analytic review. *Journal of School Psychology*, *67*, 88-103. https://doi.org/10.1016/j.jsp.2017.10.002

UN-Women. (2018). *Turning promises into action: Gender equality in the 2030 agenda for sustainable development.* United Nations. https://www.unwomen.org//media/headquarters/attachments/sections/library/publications/2018/sdg-reportgender-equality-in-the-2030-agenda-for-sustainable-development-2018en.pdf?la=en&vs=4332

*Vervecken, D., Hannover, B., & Wolter, I. (2013). Changing (s)expectations: How gender fair job descriptions impact children's perceptions and interest regarding traditionally male occupations. *Journal of Vocational Behavior*, *82*(3), 208-220. https://doi.org/10.1016/j.jvb.2013.01.008

Warren, A. S., Goldsmith, K. A., & Rimes, K.A. (2019). Childhood gender-typed behavior and emotional or peer problems: a prospective birth-cohort study. *Journal of Child Psychology and Psychiatry*, *60*(8), 888-896. https://doi:10.1111/jcpp.13051

West, S. L., Gartlehner, G., Mansfield, A. J., Poole, C., Tant, E., Lenfestey, N., Lux, L. J., Amoozegar, J., Morton, S. C., Carey, T. C., Viswanathan, M., & Lohr, K. N. (2010). *Comparative effectiveness review methods: Clinical heterogeneity*. Agency for Healthcare Research and Quality. https://www.ncbi.nlm.nih.gov/books/NBK53317/table/ch3.t2/

*Whitfield, E. L. (1978). *The effect of a planned curriculum on children's perceptions of the role of women in selected careers.* [Doctoral dissertation, Texas Tech University].

Wilson, D. B. (n.d.). *Standardized mean difference*. Practical Meta-Analysis Effect Size Calculator. https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD-main.php

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *71*(1), 136-149. https://doi.org/10.1037/0022-006X.71.1.136

World Health Organization. (n.d.). *Gender and health*. Retrieved June 7, 2021, from https://www.who.int/health topics/gender#tab=tab_1

Woolley, H. T. (1910). A review of the recent literature on the psychology of sex. *Psychological Bulletin, 7*(10), 335–342. https://psycnet.apa.org/doi/10.1037/h0066338

Younger, B. A. & Fearing, D. D. (1999). Parsing items into separate categories: Developmental change in infant categorization. *Child Development*, *70*(2), 291-303. http://dx.doi.org.libaccess.sjlibrary.org/10.1111/1467-8624.00022

Yunger, J. L., Carver, P. R., & Perry, D. G. (2004). Does gender identity influence children's psychological well-being? *Developmental Psychology*, *40*(4), 572–582. https://doi.org.libaccess.sjlibrary.org/10.1037/0012-1649.40.4.572

*Zhao, F., Zhang, Y., Alterman, V., Zhang, B., & Yu, G. (2018). Can math-gender stereotypes be reduced? A theory-based intervention program with adolescent girls. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, *37*(3), 612–624. https://doi.org/10.1007/s12144-016-9543-y

Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for sex-typed play. *Developmental Psychology*, *45*(3), 688–701. http://dx.doi.org.libaccess.sjlibrary.org/10.1037/a0014053

Zosuls, K. M., Martin, C. L., Ruble, D. N., Miller, C. F., Gaertner, B. M., England, D. E., & Hill, A. P. (2011). 'It's not that we hate you': Understanding children's gender attitudes and expectancies about peer relationships. *British Journal of Developmental Psychology*, *29*(2), 288–304. https://doi.org.libaccess.sjlibrary.org/10.1111/j.2044.835X.2010.02023.x

Appendix A

Database Search Syntaxes

ERIC 2/26/2021

(DE gender bias or sex bias or sexism or sex prejudice or gender discrimination or DE gender issues or DE sex fairness or DE sex role or DE sex stereotypes or DE sexual identity or social bias or social attitudes)

AND (DE intervention or experiment or quantitative or empirical or cognitive restructuring or positive behavior supports or psychoeducational methods or consciousness raising or attitude change)

AND (childhood attitudes or adolescent attitudes or social attitudes or children or preadolescents or high school students or junior high school students or middle school students)

AND (reduc* OR chang* decreas* OR increase* OR train OR training OR modif* OR malleab* OR moderat* or influenc*)

NOT (SU whites or SU homosexuality or SU christianity or SU teaching (occupation) or SU suicide or SU terrorism or SU Jews or SU muslims or SU political issues or SU higher education or SU impair* or SU immigrants or SU mental retardation or SU disease or SU alcohol* or SU racial relations or SU american indian reservations or SU religion or SU spouses or SU Acquired Immunodeficiency Syndrome (AIDS) or SU poverty or SU marital satisfaction or SU homeless people or SU sexual orientation or SU cancer or SU war or SU mental disorders or SU behavior disorders or Food or SU human body or SU undergraduate students or SU Agricultural Production or SU eating habits or SU college students or SU Institutionalized Persons or SU Correctional Institutions or SU death or SU refugees or SU racial attitudes or SU political attitudes or substance abuse or SU marijuana or SU Deferred Action for Childhood Arrivals (DACA) or SU Undocumented immigrants or SU COVID-19 or SU smoking or SU drug abuse or SU head injuries or SU surgery or DE racial discrimination or DE racial bias or SU intellectual disability or SU attention deficit hyperactivity disorder or SU autism or SU developmental disabilities or SU students with disabilities or SU disabilities or SU obesity)

PsycINFO 2/25/2021

(DE gender prejudice or DE sexism or DE gender bias or DE sex bias or DE gender attitudes or DE sex role attitudes or DE transgender attitudes toward or DE gender egalitarian or DE gender equality or DE gender nonconforming or DE sex discrimination or DE gender discrimination)

AND (interventions or school based intervention or prevention or prejudice reduction or flexible attitudes or attitude change)

NOT AG (adulthood or young adulthood or thirties or middle age or aged or very old)

AND (reduc* or chang* or train or training or increase or increasing or increases or increased or decrease or decreasing or decreased or decreased or malleab* or moderat* or influenc*)