



Examining the impact of fiction literature on children's gender stereotypes

Ellen E. Kneeskern^{1,2} · Patricia A. Reeder¹

Published online: 5 March 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Fiction literature has been shown to influence the beliefs and attitudes of readers of all ages. Therefore, fiction literature might be a useful tool for helping individuals overcome potentially damaging beliefs, such as rigid gender stereotypes. Since most children's literature is centered around stereotypical male protagonists, this study examined whether exposure to children's fiction literature about non-stereotypical protagonists would shift endorsement of gender stereotypes in 8- to 12-year-old children. Two experiments exposed children to book material describing the adventures of either a stereotypically or counter-stereotypically behaving male or female child protagonist. Results showed that extended exposure to a multi-chapter story about a counter-stereotypical protagonist could reduce children's endorsement of gender stereotypes; this effect was particularly strong in male children exposed to an atypical male protagonist. These findings support the importance of children's literature about protagonists who move beyond stereotypes.

Keywords Fiction literature · Gender stereotypes · Child development · Intervention

Within children's fiction literature, traditional stereotypical male protagonists are highly overrepresented (e.g., Ferguson 2018; McCabe et al. 2011). Female characters appear at less frequent rates than male characters, speak less, and have less exciting roles, with males often taking the lead in adventures. The same discrepancy even exists across children's books featuring animals as the main characters (McCabe et al. 2011; Goss 1996). Along with a general lack of female representation, many protagonists in children's literature fill stereotypical gender roles, with female characters typically taking care of children, and male characters going to work outside of the home or portraying dominant roles (e.g., Adams et al. 2011; Ferguson 2018; McCabe et al. 2011). With such biased gender representations dominating children's literature, it is

concerning to think that the books children read could reinforce children's beliefs about potentially negative gender stereotypes, particularly since research has shown that fiction literature can influence moral and empathetic development and change a reader's previously held beliefs (e.g., Ravenscroft 2012). At the same time, however, it follows that fiction literature could be used as an intervention to help children overcome any previously-held beliefs about negative gender stereotypes. This study will examine the impact of fiction chapter book content on children's perception of gender roles and characteristics, and the potential for children's literature to shift rigid gender stereotyped beliefs.

The Development and Impact of Gender Stereotypes

Gender stereotypes are the overgeneralization of certain characteristics of a group of people based entirely on that group's gender (e.g., Ellemers 2018; Flerx et al. 1976). These stereotypes can be applied to the roles individuals play in many different settings, and the characteristics and traits used to describe an individual. For example, a stereotypical female in American culture might be construed as a woman who is dependent, has long hair, and works as a secretary (Blair et al. 2001).

✉ Ellen E. Kneeskern
ekneeske@gustavus.edu

Patricia A. Reeder
preeder@gustavus.edu

¹ Department of Psychological Science, Gustavus Adolphus College, St. Peter, MN 56082, USA

² Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, ML10006, Cincinnati, OH 45229, USA

Knowledge of gender stereotypes develops early in childhood, and children as young as two years of age are aware of culturally-defined gender roles (Wilbourn and Kee 2010). One cognitive-developmental theory of gender stereotype development suggests that children progress through prescribed stages of gender knowledge and beliefs (Barth et al. 2018; Signorella et al. 1993; Trautner et al. 2005). Prior to truly forming culturally-based stereotypes about gender and sex, young infants and toddlers have the ability to discriminate between sexes based on physical characteristics and clear perceptual markers (such as hairstyle and vocal pitch (Intons-Peterson 1988)). However, this does not correspond to children having any conceptual understanding of gender or gender roles (Leinbach and Fagot 1993; Poulin-Dubois et al. 2002); they merely distinguish between genders based on easily perceived sex characteristics.

As children age, they begin to expect each gender to perform in certain ways and form associations between the types of activities and objects usually associated with men and women (Poulin-Dubois et al. 2002; Tomasetto et al. 2011; Serbin et al. 2001). On reaching pre-school age, children begin to conceptualize gender-related characteristics and activities into stereotypes about each gender. As they grow older, children then consolidate this knowledge and form rigid opinions about what each gender can do, and what it means if someone breaks the norms. For example, a child with rigid ideas about gender roles might think that if a boy puts on a dress, he automatically becomes a girl, because of the stereotype that girls wear dresses and boys do not. This thought process is typical of a child between five and seven years of age. In the third and final stage of the model, from age seven and up, children move away from the idea that gender stereotypes are strictly rigid and unchangeable, but instead start to recognize that socially-determined gender roles and characteristics can be flexible. This realization may occur due to a number of different factors, such as increased social environment flexibility based on perceptions of peer and parental behaviors, the influence of opposite-sex media character preferences, and increased opposite-sex socialization (Katz and Ksarsnak 1994).

While gender stereotypes are not inherently bad, it has been suggested that rigid and inflexible gender stereotype beliefs can have negative impacts on children's development (Aina and Cameron 2011; Cvencek et al. 2011; Kiefer and Sekaquaptewa 2007; Peterson and Lach 1990; Wilbourn and Kee 2010), and that the impact of believing in rigid gender stereotypes during childhood may impact individuals throughout their lifespan, particularly with regard to their social and work lives (Rudman and Phelan 2010; Schmid Mast 2004). For example, Cvencek et al. (2011) found that in an examination of math gender stereotypes, in both explicit and implicit measures, second grade boys and girls endorsed the stereotype that math is for boys and not girls, and girls identified as liking

math less than other subjects like reading. (This study also showed that children held math-related gender stereotype beliefs at young ages where gender-related differences in math achievement do not exist.) Importantly, similar results have been seen in adults, with female college students who identified less with math showing more negative math attitudes and lower math test scores (Nosek et al. 2002). Along with this, Rudman and Phelan (2010) showed that when women were primed with descriptions of females or males pursuing gender-stereotyped traditional careers (e.g., women becoming nurses, men becoming doctors), they had less motivation to pursue a gender-counter-stereotypical job themselves, demonstrating that simply exposing women to gender stereotypes can result in a lack of desire to endorse gender counter-stereotypic roles and characteristics.

Taken together, these results suggest that pervasive gender-based stereotypes can have potentially harmful effects on later educational and career achievement. Based on the potentially harmful effects of pervasive gender stereotypes, it is important to examine how these stereotypes could be overcome or lessened, so that their negative impacts can be reduced before they become rigid, internalized, and potentially problematic.

Fiction Literature as a Mechanism of Change

One possible mechanism for shifting beliefs and attitudes is fiction literature. The idea that fiction can enhance adults' capacity for empathy has been explored by philosophers, authors, and psychologists alike, from Aristotle to Charles Dickens to Jemeljan Hakemulder (Ravenscroft 2012). Research has shown that fiction literature can influence the moral development of adults by serving as a "moral laboratory" of sorts, wherein readers can practice empathy skills in a judgement- and consequence-free environment while learning to see things from another's point of view (Hakemulder 2000). Fiction texts also seem to be more impactful than non-fiction texts for some types of moral and conceptual development. For example, in a study of how reading material could impact multicultural acceptance, Hakemulder (2001) had university students read either a chapter of a fictional account about the difficulties of life as a woman in Algeria taken from a novel on the topic, or a nonfiction essay on general lack of rights for women in Algeria. Participants who read the fictional account of the harsh realities of life for women in Algeria were more critical of the lack of women's rights and of Algerian cultural norms for women than participants who read the nonfiction essay. This demonstrates the ability of fiction literature to increase empathy in the context of morality in ways that non-fiction accounts (which relay the same factual information) cannot.

The impact of fiction literature on readers' beliefs, attitudes, and cognition has also been studied in relation to theory

of mind mechanisms. Theory of mind is a process whereby we are able to imagine ourselves in the mind of others, capable of perceiving the beliefs, emotions, and desires of other individuals (e.g., Astington et al. 1988; Carlson et al. 2013). When we read fiction, we are tasked with using the same theory of mind mechanisms we would use in a real-life situation, in order to understand the social and emotional aspects of that situation (Kidd and Castano 2013; Oatley 2008). Just as observing a person experience a real-life situation can result in an empathetic emotional response, so can reading about a fictional character in a specific situation, due to the capacity for the same cognitive mechanisms to be engaged in each situation, real or fictional (Mar et al. 2008). These impacts of fiction literature are potentially long-lasting, as lifetime exposure to narrative fiction has been shown to have a positive association with social abilities (Mar et al. 2006). Neuroimaging research has also shown that narrative comprehension relies on four of the same areas of the brain (the medial prefrontal cortex, temporoparietal junction/posterior superior temporal sulcus, posterior cingulate, and temporal poles) that have been implicated in studies of social processing (e.g. Lieberman 2007; Mar 2004; Saxe and Wexler 2005). This suggests similar underlying neural mechanisms for both processes.

Similar to how short stories and chapter book excerpts impact adult beliefs, children's fiction picture books have been shown to influence young children's beliefs about gender stereotypes, encourage critical thinking, and teach complex concepts. Particularly when book material is paired with explicit activities that highlight the important takeaways from the book, exposure to picture books can help four- to six-year-olds shift gender attitudes. For example, when completing comprehension activities that accompanied a picture book depicting individuals in counter-stereotypic gender roles, children shifted their gender role attitudes to be more equal and were more likely to rate stereotypical male and female occupations as being appropriate for any gender (Trepanier-Street and Romatowski 1999).

In one of the most classically referenced studies on using literature to shift children's perspectives, Flerx et al. (1976) found that the presentation of egalitarian sex roles to preschoolers and kindergarteners reduced gender role stereotyping when such gender roles were presented in picture books to the children. Additionally, they found that five-year-olds were more likely to show reduced stereotypical thinking than four-year-olds, and that girls were more likely to attribute egalitarian sex roles to any gender than boys were. Along with this, Green et al. (2004) found that by presenting counter-stereotypic stories to preschool children who exhibited highly gender stereotypical play behavior, children switched to playing more neutrally rather than following strict gender stereotyped norms in their play. Likewise, Kim (2016) demonstrated that Korean-English bilingual children experienced a shift to more flexible beliefs about gender roles when they

participated in a program pairing gender-themed picture books with bilingual discussions about gender. Through this, we can see that engagement with picture books depicting counter-stereotypic information can cause children to undergo shifts in their rigid gender stereotype beliefs.

Children's fiction picture books have also been shown to teach complex intellectual concepts. In one recent study, a picture book was created and used to teach children as young as five about adaptation and evolution (Kelemen et al. 2014). The results showed that engagement with the fiction picture book helped young children learn complex concepts like adaptation, counteracting the effects of early cognitive biases against evolution. In this case, a story-book intervention program was enough to influence thinking about concepts that were thought to be unteachable to children at this age. The initial learning also endured at a three-month post-test, providing evidence that reading material can cause long-lasting impacts on child cognition.

While children's fiction picture books have demonstrated the ability to overcome rigid gender stereotypes in children as well as teach complex concepts, the content of most American children's fiction literature read at home and at school displays significant gender biases and endorses classic gender-based stereotypes. A number of content analyses of children's books published in the United States and Europe have shown that stereotypical male characters are represented more often than female characters (e.g., Crisp and Hiller 2011; Ferguson 2018; Ferguson 2019; Filipović 2018; McCabe et al. 2011). When female characters are represented in children's fiction books, they speak much less than male characters (even when they are positioned as main characters) and are likely to embody classic stereotypes about females (Kortenhaus and Demarest 1993; McCabe et al. 2011; Weitzman et al. 1972). Disappointingly, the overrepresentation of male characters in children's literature does not lead to diversity in male character portrayals: male characters are unlikely to step outside of male gender norms, rarely embodying roles such as nurturing parental figures, and are significantly more likely to serve as villains (e.g., Ferguson 2018; Ferguson 2019). Representation of transgender, nonbinary, and gender expansive characters in children's literature is similarly rare (e.g., Crawley 2017), and books that do portray counter-stereotypical characters rarely end up prominently featured in children's classrooms or libraries (e.g., Crisp et al. 2016). Given this unbalanced representation of gender systems in children's literature, it is unsurprising that even book award winners show large disparities in terms of gender stereotypes (McCabe et al. 2011). Equally distressing is the lack of awareness in some early childhood educators about these gender-based disparities in their classroom reading materials and the importance of gender representation in children's books (Filipović 2018).

Only in recent children’s literature have atypical protagonists (i.e., protagonists who are not stereotypical males) emerged in primary roles, or even in roles that were traditionally occupied by males (such as rescuer or adventurer, even though overall it is still highly unlikely that a central character is female or gender nonconforming (Goss 1996; Steyer 2014). Examination of such characters shows a trend of non-stereotypical protagonists being “exceptional” more often than not. For example, when thinking of a book like *The Hunger Games* (Collins 2008), the main protagonist is indeed female, but she is the exception to the norm of females in the society she lives in. She displays some gender stereotyped characteristics, but her counter-stereotypic behaviors swing wildly to the extreme opposite of what would be expected in an average individual (regardless of gender identity). Past research shows that when adults are exposed to “exceptional” women (such as female doctors), adult women are actually *less* likely to aspire to such a role themselves. This may be because women occupying such atypical roles are seen as the exception to the norm, and a typical female reader may believe themselves unable to achieve such an exceptional position (Rudman and Phelan 2010). The same could be true for children in that even when children are exposed to counter-stereotypical protagonists in fiction literature, these characters could be so far from the norm that they have either no impact, or a negative impact, on children’s gender stereotype beliefs.

Research makes it clear that endorsement of harmful gender stereotypes could lead to negative developmental outcomes. It is also clear that fiction picture books, adult short stories, and excerpts of novels written for adults have the capacity to lessen gender stereotypes in readers. However, children go through a significant amount of time (particularly in school) reading child-appropriate fiction chapter books, not picture books, before transitioning to reading adult-appropriate short stories and novels. Children’s fiction chapter books have been widely ignored in research, particularly in terms of their potential to cause changes in children’s beliefs and attitudes. Picture books alone cannot be relied on to create long-lasting changes in children’s perceptions of gender stereotypes, as they are written primarily for younger children who may be incapable of shifting their gender stereotype beliefs, do not typically explore complex concepts as chapter books do, and are short in length, thus reducing the potential opportunities where a reader could deeply engage and connect with the characters involved. This study will examine the use of children’s fiction chapter books as a means for shifting rigid gender stereotypes in pre-adolescent and early adolescent children. Children at these ages are capable of experiencing shifts in their beliefs about gender, and they are typically going through critical exploration of their own gender. Additionally, they have the cognitive capacity to engage deeply with age-appropriate reading

material without the need for explicit activities to draw their attention to relevant content. This study asks:

- 1) Can children’s fiction chapter books elicit the implicit reduction of gender stereotype beliefs in 8- to 12-year-old children?
- 2) How do children’s characterizations of males and females change based on their experiences with chapter book content?

Two experiments were conducted to answer these questions. The goal of Experiment 1 was to determine the impact on participants’ gender stereotypes of a single short-term exposure to an excerpt of age-appropriate fiction book material. The goal of Experiment 2 was to determine the impact of exposure to an entire book across multiple sessions on shifting rigid gender stereotypes.

Experiment 1. Methods

Participants

The participants in Experiment 1 were 8–12-year-old children from the Midwest United States ($N = 29$, $M_{age} = 9.5$ years, $SD = 1.18$ years, 58% self-identified as female). All participants had previously read at least one chapter book either by themselves, with a parent, sibling, or friend, or at school. All participants were native English speakers. Prior to obtaining child assent for participating in the study, informed consent was acquired from the parent or legal guardian of each child participant; however, neither the guardian nor the participant knew that the study was looking at gender stereotypes. Instead, they were told that the study aimed to determine children’s opinions of protagonists in children’s fiction chapter books.

Stimulus Materials

Base Text Stimulus materials were adapted from an age-appropriate, published, children’s fiction chapter book with a female child protagonist, *From The Mixed-Up Files of Mrs. Basil E. Frankweiler* (Konigsburg 1967). This book was chosen based on several specific criteria. First, it was important to find a book that participants would not have been exposed to before, so that their opinions would be based on their current exposure. Both middle school teachers and public librarians confirmed *From The Mixed-Up Files of Mrs. Basil E. Frankweiler* was unlikely to have been read by participants given popular trends in children’s literature. Likewise, the book could not have been made into a recent film, as this would make it more likely that some participants would have been exposed to the book’s content and characters prior to the

study manipulation. Lastly, the book needed to have a storyline that would be appealing, interesting, and engaging to any gender, and not written in a way that might target a specific gendered audience. This meant, for example, that the book could not have been about princesses, which would make it more likely to be written in a stereotypical fashion aimed at girls. This was done in order to make sure that the text would elicit responses based on the protagonist's gender, to make the protagonist's actions realistic and believable to the participants regardless of their or the protagonist's gender, and to eliminate possible gender-based preferences based solely on culturally-defined biases regarding the plot of the book.

Alterations to Base Text A number of edits were made to the full base text in order to create four versions of the story that could be used to test the different levels of the two independent variables in Experiments 1 and 2. First, the length of the book was cut in order to make it readable out-loud in the span of four approximately half-hour sessions. Though the base text length was cut significantly, efforts were made to retain as much information important to the primary storyline as possible, so as to not impact the rhythm of the text or the development of the main character. Only episodes in the text not relating to the overarching storyline, or areas that solely contributed to the development of secondary characters, were excluded from the final text.

After the base text was abridged, a classification process began to identify and code actions, characteristics, and traits as being stereotypically male or female. This classification process was crucial to identifying components of the text that could be flipped in order to create a stereotypical or atypical protagonist, and thus underwent three rounds of review: an initial identification of potential stereotypical components of by the lead author, then a discussion between research assistants and the lead author where the initial classifications were scrutinized and additional stereotypes were identified, and then a final review by the lead author to determine the final coding for each identified stereotype. In order to be classified, a stereotypical action, characteristic, or trait had to be related to the main character, either through the character's own descriptions of herself, descriptions by the narrator of her actions, or descriptions of her by other characters. Additionally, all stereotypical actions, characteristics, and traits, male or female, were classified based on previous research on gender stereotypes in the United States (e.g., Taylor 2003). For example, characteristics were classified as male if they involved behaviors that were dominant, independent, or intelligent (or that the protagonist was labeled as such by another character or the narrator). Characteristics were classified as female if they involved behaviors that were submissive, emotional, or receptive. (Though gender falls along a spectrum (Richards et al. 2016), this study relied on identifying and coding the most highly stereotyped actions and

characteristics that would clearly be classified as stereotypically-male or stereotypically-female based on previous research on United States cultural norms. The primary reason for focusing on these two genders and their most extreme stereotypes was to create conditions that would be as different from each other as possible while using gender pronouns and stereotypes that most children are familiar with.)

During the initial and second rounds of classification review, potential stereotypes in the text were coded as male, female, or ambiguous. The ambiguous stereotype classifications had two levels to further examine 1) Is this stereotype more male or female, and 2) Is this truly a stereotype that needs to be classified? In order to determine the classification of the ambiguous stereotypes during the second round of stereotype classification, 11 research assistants, naïve to the purpose of the study, provided ratings for these ambiguous traits, and discussions were held between raters and the lead author to determine a final code of male, female, or unnecessary. For example, in the sentence “[Character 1] informed [Character 2] that they should take advantage of the wonderful opportunity they had to learn and to study,” there is the potential argument that one character “informing” another character is a masculine stereotype. Likewise, observing that an opportunity is “wonderful” might be coded as feminine (e.g., Taylor 2003). Nine actions and traits that were initially classified as ambiguous during the first round of review were determined to be unnecessary (i.e., having “spirit” was determined to not be a stereotypical feature given the context of the word in the text), and were removed as potential targets for alteration across condition. After 100% group consensus was reached on any remaining ambiguous stereotypes, and after the final round of review by the lead author, the resulting text had 154 identified stereotypes, with 80 being classified as female and 74 being classified as male.

Next, four separate conditions (and thus four separate books) were created by editing the protagonist gender of the abridged base text and the previously-identified stereotyped characteristics: Gender Counter-Stereotypical Female (GCSF), Gender Stereotypical Female (GSF), Gender Counter-Stereotypical Male (GCSM), and Gender Stereotypical Male (GSM), based on the independent variables of protagonist gender (male or female) and the traits and characteristics the protagonist displayed (stereotypical or counter-stereotypical). The GCSF protagonist acted and appeared differently than the GSF protagonist did, and the GCSM protagonist acted and appeared differently than the GSM protagonist did. The characters were, however, paralleled, in that the GCSF and GSM versions of the protagonist were exactly the same, with the only difference being the name and gender pronouns used by the protagonist. The conditions were such that if a stereotypical action, characteristic, or trait had been classified in the coding process as male, and the condition was GCSF, that stereotype would be attributed

to the female protagonist. The same was true of the GCSM and GSF books, where these two conditions exposed participants to the same version of a protagonist displaying stereotypically female traits, with the GCSM book describing a protagonist who was identified as male and the GSF book describing a protagonist who was identified as female.

For example, if a participant was in the GCSF condition, the protagonist would be identified as female, and one of traits displayed by the protagonist would be physical aggression (a trait classified as stereotypically male by Taylor (2003) and our coding scheme). However, if the participant was in the GSF condition, the physical aggression characteristic would not be attributed to the protagonist, and was either deleted from the dialog, allocated to a secondary character, or changed to a different characteristic that fit both the storyline and the GSF condition (e.g., in the GSM or GCSF conditions a sentence might read “Claude/Claudia lost all patience, lunging towards Jamie”, whereas in the GCSM or GSF conditions it would read “Claude/Claudia lost all patience, shrieking at Jamie”). Importantly, a participant in the GSM condition would encounter a protagonist who was exactly the same as the GCSF protagonist, except for the protagonist’s gender identity. This parallel structure was used in order to control as many differences as possible across conditions.

In addition to flipping stereotyped characteristics to align with the condition type, other gendered language was modified to align with the protagonist’s gender based on condition (in addition to protagonist pronouns, gendered words like “brother” or “sister”, “boy” or “girl”, and any other gendered language directed at the protagonist were modified to fit the protagonist’s gender identity). Thus, the only altered language between the paired GCSF-GSM and GSF-GCSM conditions was this type of gendered language.

The four books that made up the stimulus materials for the four experimental conditions were reviewed by research assistants to ensure that the plot did not lose clarity with the alterations. Research assistants were also asked to classify the main character as either a GCSF, GSF, GCSM, or GSM in order to be sure that the protagonist was able to be classified by individuals not previously exposed to the material. Modifications were made to the text by the lead author until all reviewers reached 100% agreement with the condition classifications and none reported comprehension or style issues with the final versions of the stimulus materials. The final version of the text for each condition was approximately 65 single-spaced, typed pages, and the stimulus materials included no explicit mention of gender or gender roles in any way other than the pronouns and names assigned to the characters in the book. Instead, we aimed to implicitly alter gender-stereotyped attitudes exclusively through exposure to a main character who either did, or did not, occupy gender stereotypical roles in an American cultural context.

For the purposes of Experiment 1, only the first chapter of the adapted text was read to the participants. This chapter was chosen as it introduced the protagonist and plot, and had a significant number of stereotypical actions, characteristics, and traits which could clearly delineate the protagonist as being male or female and stereotypical or counter-stereotypical. Chapter 1 was edited additionally for length so that it could be read in its entirety in approximately 20 min due to time constraints. (In Experiment 2, participants were exposed to the full book for their respective condition.)

Measures

Gender-Stereotyped Attitude Scale for Children (GASC) The 13-item GASC (Signorella and Liben 1985) gives children occupations, actions, and roles, and asks them to answer “Who can” statements, such as “Who can be a doctor?” Children have the option to say either “a man,” “a woman,” or “both men and women.” The GASC includes questions about stereotypically male-oriented occupations and roles, stereotypically female-oriented occupations and roles and neutral occupations and roles (where either a male or female would be equally likely to do the stated action). This instrument is intended to measure stereotyped attitudes about gender roles, and thus was used to see if exposure to the stimulus material would impact children’s gender stereotype beliefs based on whether the stimulus material protagonist aligned with cultural stereotypes or ran counter to those stereotypes. Based on the difficulty some pilot participants had with using the GASC, we altered the survey to assist with question comprehension. Pictures of four children, two boys and two girls, and specific child model names were added to the survey answers to help participants (especially younger children). This also ensured that the child names used on the GASC were understood as male and female names. The original GASC’s response options were “Men”, “Women” or “Both men and women.” Our modified survey utilized the fictional names of the pictured children, and the response options became “Sarah,” “Jack,” or “Both Sarah and Jack” on the pre-survey, and “Claire,” “Thomas,” or “Both Claire and Thomas” on the post-survey. “Who can...” questions were also switched to “Who is more likely to...” questions. This wording seemed more likely to elicit subtle beliefs about gender stereotypes. While it is true that males or females might actually be more likely to do something (such as females being more likely to be a teacher according to current United States career statistics), the GASC test items are split in their questioning of American stereotypical occupations (which have known base rates) and stereotypical actions (which are less likely to have widely known base rates, i.e. “who is more likely to go fishing”). We anticipated that the diversity of question types would dissuade children from answering solely based on any pre-existing knowledge of gender-based base-rates. Additionally, the

questions in our version of the GASC were about child models who were not yet in the stage of life where they would actually inhabit the role of teacher, doctor, and so on. The wording “Who is more likely to...” emphasizes that we are asking about potential future roles and behaviors of the child model (rather than the model’s current roles and behaviors), allowing us to probe whether participants believe male and female models have equal likelihood of growing up to be or do anything. Along with this, the instrument includes neutral filler questions (i.e., “Who is more likely to like to do things outside?”), diversifying the overall makeup of the questions and further dissuading participants from answering non-filler questions based on prior knowledge of actual gender-based differences.

The GASC scores were coded according to how many questions were answered in a counter-stereotypic or neutral way. This was determined by first removing the responses from the three neutral filler questions. Then the number of times a participant answered “both” to a question was combined with the number of times the participant gave a non-stereotyped answer. For example, if in response to the question “Who is more likely to clean up the house?” a participant answered “Both Claire and Thomas” (the non-stereotyped answer) *or* “Thomas” (the counter-stereotypical answer), the participant would get one point added to their score. If they answered “Claire” to that same question, they would not get a point, since that was the stereotypical answer. The minimum score a participant could get was zero (showing highly stereotyped categorizations) and the maximum score was ten (showing low stereotyped categorizations). The original GASC (and similar scales such as the COAT (Liben and Bigler 2002)) was scored in a similar manner, except typically the final score was determined solely by counting the number of “both” responses. For this study, participants were given a point if they gave the non-stereotyped “both” answer *and* the counter-stereotypical answer to a question, as both responses represent a flexible attitude towards the gender stereotype being examined. Internal consistency for Experiment 1 was acceptable ($\alpha = .75$).

Reading Comprehension Test A six-item reading comprehension test was developed for the purpose of this study in order to test participant comprehension and attention to the stimulus text. Six simple questions about the plot of the book were asked in multiple choice format (i.e., “Who was the main character in this chapter?”). The questions were reviewed by elementary school teachers and professors in the English and Psychology departments at the researchers’ university in order to ensure that the questions would be able to appropriately measure comprehension and attention across participant age range. Participants who got over 50% of the test questions wrong were excluded from data analysis. Three

participants were excluded from analysis in Experiment 1 due to lack of story comprehension.

Procedure

Participants were randomly assigned to one of the four conditions (GCSF, GSF, GCSM, or GSM). Before arrival at the lab, caregivers were told that the purpose of the study was to explore different aspects of protagonists in children’s fiction chapter books. Upon arriving at the lab, both the participant and their caregiver were told that the participant would listen to a chapter from a children’s fiction book and would be asked some questions before and after listening to the excerpt.

Upon receiving caregiver consent and child verbal assent, the child completed a pre-survey on an iPad which included demographic questions, a literary term comprehension test, and the GASC. (The literary term comprehension test was included to ensure that participants would understand the questions in the reading comprehension test measure of the post-survey. No participants across either experiment were excluded due to failure to understand the literary terms used in this test, thus these data are not included in subsequent analyses.) After completing the pre-survey, participants listened to the experimenter read chapter one of the stimulus materials out loud to them. The experimenter provided a brief introduction of the material before reading, telling participants that it was about two siblings who decide to go on an adventure by running away from home and living in the Metropolitan Museum of Art in New York City for a week. After the adapted first chapter was read to the child, they were asked to complete a post-survey. The components of the post-survey were identical to the pre-survey in most ways, except it excluded demographic questions, and included the reading material comprehension test and an assessment of the main character (note that the specific questions on our outcome measurement, the GASC, was different between the pre- and post-survey, but the same number of GASC test items were present in the pre- and post-survey). After completing the post-survey, the participant returned to their caregiver, who was told that the study had been looking at how the different physical, psychological, and social aspects of book protagonists can impact children’s opinions on such protagonists.

Experiment 1 Results

To investigate whether there were significant differences in GASC scores between pre- and post-test across conditions and children, the scores were analyzed using a 2 (self-identified child gender) \times 2 (pre-to-post) \times 4 (condition) mixed ANOVA performed in R (Lawrence 2013). Level of significance was defined as $p < .05$. This analysis revealed a significant main effect of test time ($F(1, 21) = 9.45, p = .005$,

$\eta^2_G = .06$) indicating a significant difference in attitudes between the pre- ($M = 2.34, SD = 2.02$) and post-test ($M = 3.38, SD = 2.01$) GASC scores. There was no main effect of condition ($F(3, 21) = .86, p > .47, \eta^2_G = .09$), nor of child gender ($F(1, 21) = .20, p > .66, \eta^2_G = .01$). There were no significant interactions (all $p > .05$). See Table 1 and Fig. 1.

It is possible that book content may have impacted child responses to the neutral filler questions on the GASC. However, there was not a significant difference across conditions of answering more stereotypically in the GSM or GSF conditions (saying either of the children pictured were more likely versus saying both of the children pictured were more likely) or answering more neutrally in the GCSM or GCSF conditions.

Experiment 1 Discussion

Experiment 1 demonstrated that a single exposure to one chapter of a children’s fiction chapter book did not cause significant shifts in children’s beliefs about gender stereotypes, regardless of whether the participants were exposed to a stereotypical or counter-stereotypical protagonist, or a male or female protagonist. However, there was a significant main effect of test time (pre- vs. post-test) for all four conditions across the GASC, indicating that something besides the gender and characteristics of the book’s protagonist shifted children towards more gender-neutral beliefs in the post-test. Possible reasons for this are discussed more fully in the General Discussion.

There are a number of potential reasons why exposure to one chapter of the book did not elicit significant changes in gender stereotype beliefs. Despite the previously-published success of short picture books to implicitly cause changes in

children’s beliefs (e.g. Flerx et al. 1976), it is possible that children in this age range need more exposure to book content in order to experience shifts in beliefs. It is also possible that a single chapter of a multi-chapter book fails to engage a reader with enough in-depth material about the characters to allow the reader to encode how much the protagonist aligned with gender stereotypes. To test these possibilities, a second experiment was conducted in order to determine the impact that multiple exposures to counter-stereotypical protagonists might have on pre-adolescents.

Experiment 2. Methods

Participants

Participants were 50 fifth and sixth grade students in three classrooms in two public middle schools in the Midwest United States (School A: $N = 16; M_{age} = 11, SD = .61, 37\%$ self-identified females; School B: $N = 34; M_{age} = 11.3, SD = .46, 55\%$ self-identified females). All participants had read at least one chapter book previously either by themselves, with a parent, sibling, or friend, or at school, and all participants were native English speakers, with one child indicating they also spoke some Italian at home, four children indicating they also spoke some Spanish at home, and one child indicating they also spoke some Japanese at home. No Experiment 1 participant also participated in Experiment 2.

Measures

The measures used in Experiment 2 were identical to the measures used in Experiment 1, except that the post-survey had different comprehension questions more appropriate for testing comprehension of the entire book. Additionally, after each reading session, participants in Experiment 2 were asked to write one sentence summarizing what they had heard that day in order to ensure they understood the content of the book material. Internal consistency on the GASC for Experiment 2 was good ($\alpha = .85$).

Procedure

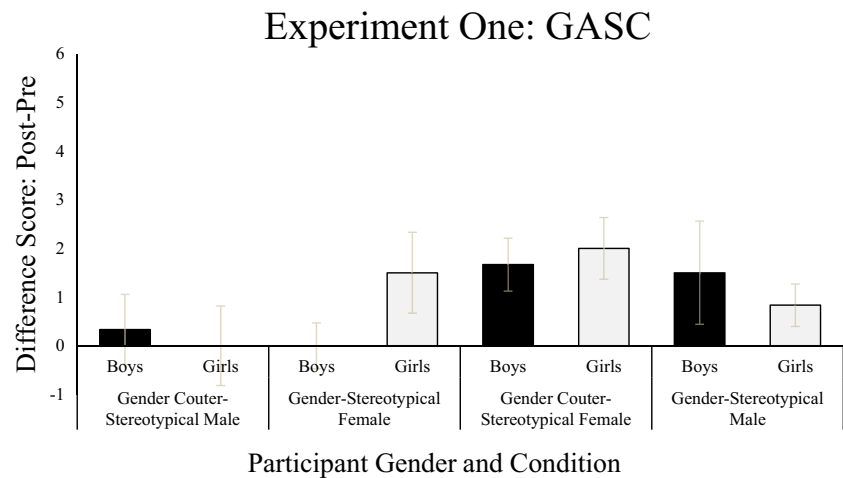
Informed site consent was acquired from the principal of each school prior to asking for parent/guardian consent, and the principals and teachers involved in Experiment 2 were informed of the true purpose of the study, but explicitly told not to reveal the true purpose to their students. Each classroom was randomly assigned to one of three conditions: GCSF, GSF, or GCSM. The fourth condition, GSM, was excluded due to the number of classrooms available to work with, and the high likelihood that participants had already been exposed to age-appropriate fiction with stereotypical male protagonists

Table 1 Descriptive statistics for Experiment 1

		GASC ($N = 29$, scores range 0–10)	
		Pre-test $M (SD)$	Post-test $M (SD)$
GCSM	Male	1 (.82)	1.3 (.94)
	Female	3 (2.16)	3 (1.63)
GSM	Male	2.5 (.5)	4 (2)
	Female	1.67 (1.49)	2.5 (1.5)
GCSF	Male	3 (0)	4.67 (.94)
	Female	2.6 (3.72)	4.6 (3)
GSF	Male	2 (.63)	3.5 (1)
	Female	3.67 (.47)	3.67 (.47)

GASC Gender Stereotyped Attitude Scale for Children, GCSM Gender Counter-Stereotypical Male, GSM Gender Stereotypical Male, GCSF Gender Counter-Stereotypical Female, GSF Gender Stereotypical Female

Fig. 1 The results of the *Gender Stereotyped Attitude Scale for Children (GASC)* for participants in Experiment 1. The figure represents the average difference score of participants' post-test GASC score minus pre-test GASC score across condition and child gender. Error bars represent the standard error of the mean. Level of significance was defined as $p < .05$. No significant differences between conditions were found



in most of their prior reading material. (We might even consider the pre-survey GASC scores to serve as a baseline measure of stereotype beliefs after regular exposure to GSM protagonists, given their ubiquity in children's literature.) The book materials and dependent measures used in Experiment 2 were identical to those used in Experiment 1, except that the full text of each book condition was used for the purposes of Experiment 2, and the post-survey had different comprehension questions more appropriate for testing comprehension of the entire book. Additionally, after each reading session, participants in Experiment 2 were asked to write one sentence summarizing what they had heard that day in order to ensure they understood the content of the book material. The exact procedure differed slightly between schools due to teacher preferences and time constraints.

School A (GCSM Condition) Participants self-selected into the study based on the description of the study given to them by the teacher assisting in the study. The researcher met with the students six times over the course of two weeks, with the first and last session entailing of administration of the pre- and post-surveys. Before the pre- and post-surveys, the students were told that the survey was completely anonymous. They were asked to answer each question truthfully and with the first thought that came to mind. Additionally, they were told that there were no wrong answers to any of the questions on the surveys. The participants were also told that the pre- and post-surveys were to be done individually and that they were not allowed to talk to the people around them while they completed each survey. All of the pre- and post-surveys were administered via paper surveys, which were later electronically coded by trained research assistants. The other four sessions were reading exposures, where the children sat and listened to sequential sections of the story read out loud by the experimenter. After each reading

session, the participants each filled out a tracked daily comprehension survey which had nothing to do with gendered attitudes or stereotypes, and simply asked participants to "Please write one sentence summarizing what happened in today's reading." Each session was approximately 30 min long. (Note that reading aloud is an accepted practice in many middle school classrooms in the United States (e.g., Albright and Ariail 2005; Ivey 2003; Marchessault and Larwin 2013), and we confirmed with teachers at School A & B that our participants had been in classrooms where teachers read aloud from texts during class. Thus, we do not believe that it was outside of our participants' normal classroom experience to listen to the experimenter read chapters aloud from a book and then ask comprehension questions about the readings.)

School B (GCSF and GSF Conditions) Participants were selected for this study by their teachers and given the same description of the study that participants in School A were given. Participants from School B also gave verbal assent to participate in the study and were told that they did not have to participate if they did not want to, with no consequences. The researcher first met with each class individually and administered the pre-survey approximately one month before the first reading exposure. The instructions given were identical to the ones given to the participants at School A. One month after the pre-survey was given, the researcher returned to each classroom individually, and read sequential sections of the book out loud to the participants for a total of four exposure sessions at School B. The only difference in the reading sessions was that the sessions were done back-to-back, with one every day for a week, rather than being conducted every other day as was done at School A. On the final day of the study, the researcher administered the post-survey to the participants with the same instructions given as at the time of the pre-survey.

Experiment 2 Results

To investigate whether there were significant differences in GASC scores between pre- and post-test across conditions and children, the scores were analyzed using a 2 (self-identified child gender) × 2 (pre-to-post) × 3 (condition) mixed ANOVA performed in R (Lawrence 2013). Level of significance was defined as $p < .05$. Two additional 2 (pre- or post-test) × 3 (condition) mixed ANOVAs were conducted examining the scores on the GASC for female and male participants separately.

The 2 (child gender) × 2 (pre-to-post) × 3 (condition) mixed ANOVA revealed a significant main effect of test time ($F(1, 44) = 25.49, p < .001, \eta^2_G = .11$), indicating a significant difference in attitudes between the pre-test ($M = 3.78, SD = 2.92$) and post-test ($M = 5.74, SD = 3.22$) scores. There was also a significant main effect of participant gender ($F(1, 44) = 7.62, p = .008, \eta^2_G = .12$). There were marginally significant interactions between condition and test time ($F(2, 44) = 2.98, p = .061, \eta^2_G = .02$), and between condition and participant gender ($F(2, 44) = 3.02, p = .059, \eta^2_G = .09$). An analysis of the neutral filler questions showed no significant differences across conditions (see Table 2 and Fig. 2).

In the analysis of the male participants, there was a significant interaction between condition and test time ($F(2, 22) = 4.05, p = .03, \eta^2_G = .09$), showing that the change before and after being exposed to the reading material differed significantly across conditions. The biggest difference between pre- and post-test GASC scores was seen in boys who were exposed to book content with a counter-stereotypic male protagonist ($M = 3.8, SD = 3$).

The analysis of the female participants revealed a significant main effect of test time ($F(1, 22) = 10.10, p = .004, \eta^2_G = .08$), indicating a significant difference in attitudes between the pre-test ($M = 4.92, SD = 3.24$) and post-test ($M = 6.64, SD = 3.07$) scores, regardless of condition. There was no

significant main effect of condition, and no significant interaction between condition and test time ($p > .40$ for both).

General Discussion

The results of this study suggest that children who engage with a full multi-chapter book that portrays a counter-stereotypical protagonist may experience shifts in their beliefs about gender stereotypes. Interestingly, the results of Experiment 1 suggest that a single exposure to one chapter of a multi-chapter book is not sufficient to shift gender stereotyped beliefs, in contrast to previous research that showed that exposure to a short fiction picture book could shift children’s beliefs on stereotyped gender roles (e.g., Trepanier-Street and Romatowski 1999). However, it is important to note that past studies also often combined their book exposures with activities highlighting gender, which this study did not do, choosing to look only at learning through book material that did not explicitly discuss gender or gender roles. It is also important to note that the majority of past research was on younger children under the age of seven and exposed them to fiction picture books multiple times. Thus, while each exposure session in previous research provided the same amount of time with non-stereotyped gender content as the children in Experiment 1 received (approximately 20 min), we can see that perhaps it was the multiple exposures to a shorter storyline and explicit highlighting of gender-related content that caused shifts in beliefs in previous research, along with the younger age of the participants. It is also important to consider that the illustrations in picture books could have contributed to how children’s beliefs about stereotypes may have factored into previous results, and how the lack of pictures depicting counter-stereotypic information in Experiment 1 could have contributed to the lack of change in beliefs.

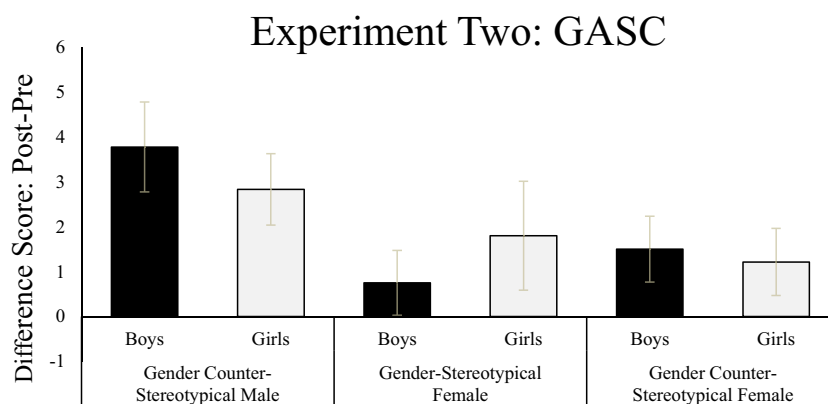
Experiment 2 demonstrated that a multiple-exposure fiction book reading program has the capacity to shift children’s beliefs about gender stereotypes. This result was strongest in male children who were exposed to a storyline involving a counter-stereotypical male protagonist. The results of Experiment 2 are important because they indicate that middle-school children act in similar ways to adult participants when exposed to fiction chapter books and have the capacity to shift their beliefs based solely on exposure to fiction literature (e.g., Hakemulder 2001). This supports the idea that reading material can implicitly impact early adolescent beliefs and attitudes. These results are also interesting in that the most altered beliefs were seen in male participants in reaction to the GCSM protagonist, which is contrary to previous studies that suggest females shift their beliefs about gender more than males do when exposed to characters that do not align with gender stereotypes (Green et al. 2004). This suggests that perhaps male children are not exposed to GCSM

Table 2 Descriptive statistics for Experiment 2

		GASC ($N = 50$, scores range 0–10)	
		Pre-test $M (SD)$	Post-test $M (SD)$
GCSM	Male	3.1 (2.74)	7 (3.29)
	Female	3.17 (1.57)	6 (3.11)
GCSF	Male	2.2 (1.13)	3 (1.41)
	Female	5.8 (3.82)	7.6 (2.33)
GSF	Male	2.5 (1.12)	4 (2.38)
	Female	5.36 (3.29)	6.57 (3.2)

GASC Gender Stereotyped Attitude Scale for Children, GCSM Gender Counter-Stereotypical Male, GCSF Gender Counter-Stereotypical Female, GSF Gender Stereotypical Female

Fig. 2 The results of the *Gender Stereotyped Attitude Scale for Children* (GASC) for participants in Experiment 2. The figure represents the average difference score of participants' post-test minus pre-test across condition and child gender. Error bars represent the standard error of the mean. There was a significant interaction between condition and test time for male participants in the study, with the largest difference seen in boys exposed to a counter-stereotypical male protagonist



Participant Gender and Condition

protagonists as often as girls are exposed to GCSF protagonists, a possibility supported by an increased focus in modern society on counter-stereotypical females in current media (in films, such as *The Hunger Games Trilogy* (Collins 2008), in social media, such as the Facebook group “A Mighty Girl,” (A Mighty Girl 2012) and potentially in literature, such as *Out of My Mind* (Draper 2010)), but a lack of representation of counter-stereotypical males (e.g., Adams et al. 2011; Crisp et al. 2016; Gordon and Roberts 2016). In addition to male children having limited exposure to GCSM protagonists in literature and media, boys could be more likely than girls to be stigmatized if seen behaving in counter-stereotypical manners or endorsing counter-stereotypic roles. In this study in particular, the counter-stereotypical male protagonist was the exact parallel to the stereotypical female protagonist and portrayed no stereotypically male traits or actions and minimal neutral traits or actions. This is likely rare in fiction literature targeted at young male readers (e.g., Gordon and Roberts 2016).

An alternative explanation for the significant shift in beliefs seen in male participants in Experiment 2 (but not female participants) is that male participants had lower initial pre-test scores than females did, particularly for the GASC measure. Female children were close to ceiling for the GASC pre-test scores (see Table 2), whereas male children were significantly more stereotyped in their beliefs. Thus, males had more opportunity to show growth through this measure than females. Regardless, it is interesting to note this asymmetry in initial GASC ratings between male and female participants, and it suggests that the increased presence of counter-stereotypical female protagonists in film, social media, and literature might push females towards holding fewer stereotyped beliefs. There is also previous literature suggesting that female children’s beliefs are more malleable than males’ in general, which could have caused the higher female pre-scores (Green et al. 2004).

It is important to note that while it is possible that mere exposure to the GASC could act as an intervention in itself, this appears not to be the case. Children in Experiment 1 did not dramatically shift their answers to be 100% counter-stereotypical in the post-survey, despite having taken the pre-survey less than an hour before, and in fact did not show significant changes in beliefs. If the GASC had acted as a primer or an additional intervention stimulus, we would expect to see the Experiment 1 post-tests show complete, or near-complete, counter-stereotypical responses. Given that participants in Experiment 2 had weeks of time between the pre- and post-tests, and that gender was never mentioned explicitly on either test, it is unlikely that anything but the designated stimulus text caused changes in participant attitudes towards gender stereotypes.

Limitations

There are some limitations in the current study that could be addressed by future research. Notably, Experiment 2 did not test the GSM condition (under the constraints of the number of classrooms available for the study, and the belief that gender-stereotypical male protagonists are exceedingly common in children’s fiction literature and serve as a baseline condition). Because of the lack of inclusion of the GSM condition, it is difficult to draw strong conclusions about results from the GSF condition alone. Moreover, due to the limitation of having to assign classrooms to conditions in Experiment 2, the self-identified genders of the participants and their teachers were unable to be controlled or matched across conditions, resulting in unbalanced numbers of male and female participants in each condition. The children assigned to the GCSF condition in Experiment 2 had a male teacher, while the other Experiment 2 participants all had female teachers; the lack of random assignment of participants to condition makes it

difficult to separate potential effects of teacher gender from participant behavior. It should be noted that book comprehension after each reading session was similar across all three groups of participants, as well as similarities in demographic measures and engagement during each reading session, suggesting that other differences between the groups are less likely to fully explain differences in GASC results.

Additionally, although the GASC is a previously published and validated instrument, it is an older measure and could be testing outdated stereotyped beliefs. While the more recently developed Children's Occupation, Activity, and Trait – Attitude Measure (COAT-AM) and Children's Occupation, Activity, and Trait – Personal Measure (COAT-PM) (Liben and Bigler 2002) scales could have been used as the primary measurement of gender stereotype beliefs, the authors found the GASC to be more appropriate for the purposes of this study and its timeline. The GASC is very similar to the COAT scales in its content, including the same occupation and activity items, but the COAT scales are much longer in both their forms. Even in the short form versions, the COAT scales include nearly double the number of items than the GASC. Due to the time constraints of testing in a classroom setting and participant attention spans, the GASC seemed to be a better choice for the study, particularly because it is still being used in recent research (e.g., Nathanson 2010). Nevertheless, it is possible that the lack of recent norming for the GASC, combined with our wording changes, explain the ceiling effects we saw in the female participants' GASC scores, making it a less useful measure for comparing results across all participants. Lastly, though the authors took care in balancing the pre- and post-test survey questions, the main effect of test time in both experiments suggests that the lack of randomization of GASC questions across participants led to some more stereotyped questions ending up on the pre-test and less stereotyped questions ending up on the post-test. In particular, although all of the stereotyped questions on the GASC were supposed to be equally stereotypical, due to the age of the instrument, this may not have been the case.

Future Directions

An important extension of this study would look at the long-term impacts of exposure to counter-stereotypical and non-conforming protagonists on adolescents and older children, and how shifts in beliefs about gender stereotypes might persist after the exposure to counter-stereotypical protagonists ends. It also would be interesting to look at how repeated exposure to non-traditional literary protagonists can shift views, and how many books about counter-stereotypical protagonists children need to be exposed to for long term changes in beliefs to occur. Future studies could also look at the impact of fiction chapter books on children who are both younger or older than the children in this study and see if chapter books

without pictures can teach young children about gender roles, examine how a non-binary protagonist could impact children's ideas about gender roles, or determine if teenagers are able to implicitly learn through young adult chapter books in the same way that adults are able to do with novels. Additionally, this study involved exposure to the book material via the experimenter reading the text out loud. How might the results of the study shift if children read the materials by themselves, at their own pace, and physically seeing some of the nuances of the text as opposed to hearing them? Finally, future research should ask if fiction literature could be impactful in shifting other stereotyped beliefs in children of this age, such as beliefs about multiculturalism, social class, or race and ethnicity. If it can be determined that fiction literature content alone can have significant impacts on the development and shifting of numerous stereotypical beliefs, as this study has begun to demonstrate with gender stereotypes, then future intervention programs could focus on literature as an effective means of helping children overcome harmful stereotypes.

Acknowledgments We would like to thank Kyle Chambers, Robert Kendrick, Leanne Tamm, Stephen Becker, and the members of the Gustavus Adolphus College Center for Developmental Science for their helpful comments on this work, with special thanks to Mark Kruger and Leanne Tamm for assistance with data analysis and interpretation. We also thank Eva Hendrickson, Andrea Viznor, and Andrew Jurek for allowing us to work with their students and facilitating our work in the schools.

Compliance with Ethical Standards

Conflict of Interest The authors do not have conflicts of interest to disclose. There are no funding sources to disclose. The study was approved by the relevant Institutional Review Board (study approval #1718–0061). Since all participants were under the age of 18, informed consent to participate in the study was obtained from the parents/guardians of all participants, and informed assent was obtained from child participants prior to participation. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- A Mighty Girl. (2012). In Facebook [Group page]. Retrieved January 22, 2019, from <https://www.facebook.com/amightygirl/>.
- Adams, M., Walker, C., & O'Connell, P. (2011). Invisible or involved fathers? A content analysis of representations of parenting in young children's picturebooks in the UK. *Sex Roles, 65*, 259–270. <https://doi.org/10.1007/s11199-011-0011-8>.
- Aina, O. E., & Cameron, P. A. (2011). Why does gender matter? Counteracting stereotypes with young children. *Dimensions of Early Childhood, 39*(3), 11–20.
- Albright, L. K., & Ariail, M. (2005). Tapping the potential of teacher read-alouds in middle schools. *Journal of Adolescent & Adult Literacy, 48*(7), 582–591.
- Astington, J. W., Harris, P. L., & Olson, D. R. (Eds.). (1988). *Developing theories of mind*. New York: Cambridge University Press.

- Barth, J. M., Kim, H., Eno, C. A., & Guadagno, R. E. (2018). Matching abilities to careers for others and self: Do gender stereotypes matter to students in advanced math and science classes? *Sex Roles*, *79*(1–2), 83–97.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828–841. <https://doi.org/10.1037/0022-3514.81.5.828>.
- Carlson, S. M., Koenig, M. A., & Harms, M. B. (2013). Theory of mind. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(4), 391–402. <https://doi.org/10.1002/wcs.1232>.
- Collins, S. (2008). *The hunger games*. New York: Scholastic Press.
- Crawley, S. A. (2017). Be who you are: Exploring representations of transgender children in picturebooks. *Journal of Children's Literature*, *43*(2), 28–41.
- Crisp, T., & Hiller, B. (2011). "Is this a boy or a girl?": Rethinking sex-role representation in Caldecott medal-winning picturebooks, 1938–2011. *Children's Literature in Education*, *42*(3), 196–212. <https://doi.org/10.1007/s10583-011-9128-1>.
- Crisp, T., Knezek, S. M., Quinn, M., Bingham, G. E., Girardeau, K., & Starks, F. (2016). What's on our bookshelves? The diversity of children's literature in early childhood classroom libraries. *Journal of Children's Literature*, *42*(2), 29–42.
- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, *82*(3), 766–779. <https://doi.org/10.1111/j.1467-8624.2010.01529.x>.
- Draper, S. (2010). *Out of my mind*. New York: Simon & Schuster.
- Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*, 275–298. <https://doi.org/10.1146/annurev-psych-122216-011719>.
- Ferguson, D. (2018). Must monsters always be male? Huge gender bias revealed in children's books. The Guardian. Retrieved from <https://www.theguardian.com/books/2018/jan/21/childrens-books-sexism-monster-in-your-kids-book-is-male>.
- Ferguson, D. (2019). 'Highly concerning': Picture books bias worsens as female characters stay silent. The Guardian. Retrieved from <https://www.theguardian.com/books/2019/jun/13/highly-concerning-picture-books-bias-worsens-as-female-characters-stay-silent>.
- Filipović, K. (2018). Gender representation in children's books: Case of an early childhood setting. *Journal of Research in Childhood Education*, *32*(3), 310–325. <https://doi.org/10.1080/02568543.2018.1464086>.
- Flerx, V. C., Fidler, D. S., & Rogers, R. W. (1976). Sex role stereotypes: Developmental aspects and early intervention. *Child Development*, *47*(4), 998–1007. <https://doi.org/10.2307/1128436>.
- Gordon, N., & Roberts, S. K. (2016). Boys will be boys?: Toward a gender-balanced literature curriculum. *Florida Reading Journal*, *51*(2), 7–17.
- Goss, G. (1996, February). *Weaving girls into the curriculum*. Chicago: Paper presented at the Annual Meeting of the Association of Colleges for Teacher Education.
- Green, V. A., Bigler, R., & Catherwood, D. (2004). The variability and flexibility of gender-typed toy play: A close look at children's behavioral responses to counterstereotypic models. *Sex Roles*, *51*, 371–386. <https://doi.org/10.1023/B:SERS.0000049227.05170.a>.
- Hakemulder, J. (2000). *The moral laboratory: Experiments examining the effects of reading literature on social perception and moral self-concept*. Amsterdam: Benjamins.
- Hakemulder, J. (2001). How to make alle Menschen Bruder: Literature in a multicultural and multiform society. In Ravenscroft, I. (2012). Fiction, imagination, and ethics. In R. Langdon, C. Mackenzie, R. Langdon, C. Mackenzie (Eds.), *Emotions, imagination, and moral reasoning* (71–89). New York: Psychology Press.
- Intons-Peterson, M. J. (1988). *Children's concepts of gender*. Norwood: Ablex.
- Ivey, G. (2003). "The teacher makes it more explainable" and other reasons to read aloud in the intermediate grades. *The Reading Teacher*, *56*(8), 812–814.
- Katz, P. A., & Ksanskak, K. R. (1994). Developmental aspects of gender role flexibility and traditionality in middle childhood and adolescence. *Developmental Psychology*, *30*(2), 272–282. <https://doi.org/10.1037/0012-1649.30.2.272>.
- Kelemen, D., Emmons, N. A., Seston Schillaci, R., & Ganea, P. A. (2014). Young children can be taught basic natural selection using a picture-storybook intervention. *Psychological Science*, *25*(4), 893–902. <https://doi.org/10.1177/0956797613516009>.
- Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, *342*(6156), 377–380. <https://doi.org/10.1126/science.1239918>.
- Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychological Science*, *18*(1), 13–18. <https://doi.org/10.1111/j.1467-9280.2007.01841.x>.
- Kim, S. J. (2016). Expanding the horizons for critical literacy in a bilingual preschool classroom: Children's responses in discussions with gender-themed picture books. *International Journal of Early Childhood*, *48*(3), 311–327. <https://doi.org/10.1007/s13158-016-0171-3>.
- Konigsburg, E. L. (1967). *From the mixed-up files of Mrs. Basil E. Frankweiler*. Atheneum Books for Young Readers.
- Kortenhaus, C. M., & Demarest, J. (1993). Gender role stereotyping in children's literature: An update. *Sex Roles*, *28*(3–4), 219–232. <https://doi.org/10.1007/bf00299282>.
- Lawrence, M. A. (2013). ez: Easy analysis and visualization of factorial experiments. Retrieved from <http://cran.rproject.org/web/packages/ez>.
- Leinbach, M. D., & Fagot, B. I. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. *Child Behavior and Development*, *16*(3), 317–332. [https://doi.org/10.1016/0163-6383\(93\)80038-a](https://doi.org/10.1016/0163-6383(93)80038-a).
- Liben, L. S., & Bigler, R. S. (2002). Introduction. *Monographs of the Society for Research in Child Development*, *67*, 1–6. <https://doi.org/10.1111/1540-5834.t01-1-00188>.
- Lieberman, M. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology*, *58*, 259–289. <https://doi.org/10.1146/annurev.psych.58.110405.085654>.
- Mar, R. A. (2004). The neuropsychology of narrative: story comprehension, story production and their interrelation. *Neuropsychologia*, *42*(10), 1414–1434. <https://doi.org/10.1016/j.neuropsychologia.2003.12.016>.
- Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality*, *40*(5), 694–712. <https://doi.org/10.1016/j.jrp.2005.08.002>.
- Mar, R., Djikie, M., & Oatley, K. (2008). Effects of reading on knowledge, social abilities, and selfhood: Theory and empirical studies. In S. Zyngier, M. Bortolussi, A. Chesnokova, & J. Auracher (Eds.), *Directions in empirical literary studies: In honor of Willie van peer* (pp. 127–137). Amsterdam: Benjamins.
- Marchessault, J. K., & Larwin, K. H. (2013). Structured read-aloud in middle school: The potential impact on reading achievement. *Contemporary Issues in Education Research*, *6*(2), 241–246.
- McCabe, J., Fairchild, E., Grauerholz, L., Pescosolido, B. A., & Tope, D. (2011). Gender in twentieth-century children's books. *Gender & Society*, *25*(2), 197–226. <https://doi.org/10.1177/0891243211398358>.
- Nathanson, A. I. (2010). Using television mediation to stimulate nontraditional gender roles among Caucasian and African American children in the US. *Journal of Children and Media*, *4*(2), 174–190.

- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, 83(1), 44–59. <https://doi.org/10.1037/0022-3514.83.1.44>.
- Oatley, K. (2008). The science of fiction. *New scientist*, (42–43). In Ravenscroft, I. (2012). Fiction, imagination, and ethics. In R. Langdon, C. Mackenzie, R. Langdon, C. Mackenzie (Eds.), *Emotions, imagination, and moral reasoning* (71–89). New York: Psychology Press.
- Peterson, S. B., & Lach, M. A. (1990). Gender stereotypes in children's books: Their prevalence and influence on cognitive and affective development. *Gender and Education*, 2(2), 185–197. <https://doi.org/10.1080/0954025900020204>.
- Poulin-Dubois, D., Serbin, L. A., Eichstedt, J. A., Sen, M. G., & Beissel, C. F. (2002). Men don't put on make-up: Toddlers' knowledge of the gender stereotyping of household activities. *Social Development*, 11(2), 166–181. <https://doi.org/10.1111/1467-9507.00193>.
- Ravenscroft, I. (2012). Fiction, imagination, and ethics. In R. Langdon, C. Mackenzie, R. Langdon, & C. Mackenzie (Eds.), *Emotions, imagination, and moral reasoning* (pp. 71–89). New York: Psychology Press.
- Richards, C., Bouman, W. P., Seal, L., Barker, M. J., Nieder, T. O., & T'Sjoen, G. (2016). Non-binary or genderqueer genders. *International Review of Psychiatry*, 28(1), 95–102. <https://doi.org/10.3109/09540261.2015.1106446>.
- Rudman, L. A., & Phelan, J. E. (2010). The effect of priming gender roles on women's implicit gender beliefs and career aspirations. *Social Psychology*, 41(3), 192–202. <https://doi.org/10.1027/1864-9335/a000027>.
- Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, 43(10), 1391–1399. <https://doi.org/10.1016/j.neuropsychologia.2005.02.013>.
- Schmid Mast, M. (2004). Men are hierarchical, women are egalitarian: An implicit gender stereotype. *Swiss Journal Of Psychology / Schweizerische Zeitschrift Für Psychologie/Revue Suisse De Psychologie*, 63(2), 107–111. <https://doi.org/10.1024/1421-0185.63.2.107>.
- Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development*, 25(1), 7–15. <https://doi.org/10.1080/01650250042000078>.
- Signorella, M. L., & Liben, L. S. (1985). Assessing children's gender-stereotyped attitudes. *Psychological Documents*, 15(7). <https://doi.org/10.1037/t09185-000>.
- Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review*, 13, 147–183. <https://doi.org/10.1006/drev.1993.1007>.
- Steyer, I. (2014). Gender representations in children's media and their influence. *Campus-Wide Information Systems*, 31(2/3), 171–180. <https://doi.org/10.1108/CWIS-11-2013-0065>.
- Taylor, F. (2003). Content analysis and gender stereotypes in children's books. *Teaching Sociology*, 31(3), 300–311. <https://doi.org/10.2307/3211327>.
- Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology*, 47(4), 943–949. <https://doi.org/10.1037/a0024047>.
- Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development*, 14(4), 365–381. <https://doi.org/10.1002/icd.399>.
- Trepanier-Street, M. L., & Romatowski, J. A. (1999). The influence of children's literature on gender role perceptions: A reexamination. *Early Childhood Education Journal*, 26(3), 155–159. <https://doi.org/10.1023/A:1022977317864>.
- Weitzman, L. J., Eifler, D., Hokada, E., & Ross, C. (1972). Sex-role socialization in picture books for preschool children. *American Journal of Sociology*, 77(6), 1125–1150. <https://doi.org/10.1086/225261>.
- Wilbourn, M. P., & Kee, D. W. (2010). Henry the nurse is a doctor too: Implicitly examining children's gender stereotypes for male and female occupational roles. *Sex Roles*, 62(9–10), 670–683. <https://doi.org/10.1007/s11199-010-9773-7>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.