Journal of School Psychology 53 (2015) 25–44

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis☆

Paulette C. Flore *, Jelte M. Wicherts

Tilburg University, The Netherlands

ARTICLE INFO

Article history:
Received 26 November 2013
Received in revised form 24 October 2014
Accepted 25 October 2014
Available online 13 November 2014

Keywords:
Stereotype threat
Math/science test performance
Gender gap
Test anxiety
Publication bias
Meta-analysis

ABSTRACT

Although the effect of stereotype threat concerning women and mathematics has been subject to various systematic reviews, none of them have been performed on the sub-population of children and adolescents. In this meta-analysis we estimated the effects of stereotype threat on performance of girls on math, science and spatial skills (MSSS) tests. Moreover, we studied publication bias and four moderators: test difficulty, presence of boys, gender equality within countries, and the type of control group that was used in the studies. We selected study samples when the study included girls, samples had a mean age below 18 years, the design was (quasi-)experimental, the stereotype threat manipulation was administered between-subjects, and the dependent variable was a MSSS test related to a gender stereotype favoring boys. To analyze the 47 effect sizes, we used random effects and mixed effects models. The estimated mean effect size equaled −0.22 and significantly differed from 0. None of the moderator variables was significant; however, there were several signs for the presence of publication bias. We conclude that publication bias might seriously distort the literature on the effects of stereotype threat among schoolgirls. We propose a large replication study to provide a less biased effect size estimate.

1. Introduction

Spencer, Steele, and Quinn (1999) first suggested that women's performance on mathematics tests could be disrupted by the presence of a *stereotype threat*. This initial paper inspired many researchers to replicate the stereotype threat effect and expand the theory by introducing numerous moderator variables and various dependent variables related to negative gender stereotypes, such as tests of Mathematics, Science, and Spatial Skills (MSSS). This practice resulted in approximately one hundred research papers and five meta-analyses (Nguyen & Ryan, 2008; Picho, Rodriguez, & Finnie, 2013; Stoet & Geary, 2012; Walton & Cohen, 2003; Walton & Spencer, 2009). Although four of these systematic reviews (Nguyen & Ryan, 2008; Picho et al., 2013; Walton & Cohen, 2003; Walton & Spencer, 2009) confirmed the existence of a robust mean stereotype threat effect, some ambiguities regarding this effect remain. For instance, it has been suggested (*Ganley et al., 2013; Stoet & Geary, 2012) that the stereotype threat literature is subject to an *excess of significant findings*, which might be caused by publication bias (Ioannidis, 2005; Rosenthal, 1979), *p-hacking* (i.e., using questionable research practices to obtain a statistically significant effect; Simonsohn, Nelson, & Simmons, 2013), or both (Bakker, van Dijk, & Wicherts, 2012). A less controversial but nevertheless interesting issue is the age at which stereotype threat begins to influence performance on MSSS tests: does stereotype threat already influence children's performance, or does this effect

☆ The preparation of this article was supported by grant numbers 016-125-385 and 406-12-137 from the Netherlands Organization for Scientific Research (NWO).
* Corresponding author at: Department of Methodology and Statistics, Tilburg School of Behavioral and Social Sciences, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.
E-mail address: P.C.Flore@tilburguniversity.edu (P.C. Flore).
Action Editor: Craig Albers

http://dx.doi.org/10.1016/j.jsp.2014.10.002
0022-4405/

only emerge during early adulthood? Both of these issues are addressed in this article by means of a meta-analysis of the stereotype threat literature in the context of schoolgirls' MSSS test performance. We will introduce these topics by providing a general review of the literature on stereotype threat and the onset of gender differences in the domains of MSSS.

1.1. Stereotype threat

The effect of stereotype threat refers to the ramifications of an activated negative stereotype or an emphasized social identity (Steele, 1997). Individuals who are members of a stigmatized group tend to perform worse on stereotype relevant tasks when confronted with that negative stereotype (Steele & Aronson, 1995). In their seminal paper, Steele and Aronson (1995) focused on ethnic minorities as stereotyped group. Later experiments showed similar effects for other stigmatized groups, including women in the quantitative domain (e.g., Ambady, Paik, Steele, Owen-Smith, & Mitchell, 2004; Brown & Josephs, 1999; Oswald & Harvey, 2001; Schmader & Johns, 2003; Spencer et al., 1999). In these experiments, women were either assigned to a stereotype threat condition, where they were exposed to a gender-related stereotype threat (e.g., a written statement that men perform better on mathematics tests than women), or to a control condition, where they were not exposed to such a threat. When participants subsequently completed a MSSS test (e.g., a mathematical test), women who were assigned to the stereotype threat condition averaged lower scores than women who were assigned to the control condition (Ambady et al., 2004; Brown & Josephs, 1999; Oswald & Harvey, 2001; Schmader & Johns, 2003; Spencer et al., 1999). The results of these studies were deemed important, because researchers suspected that stereotype threat could be a driving force behind the decision of women to leave the science, technology, engineering, and mathematics (STEM) fields (Cheryan & Plaut, 2010; Schmader, Johns, & Barquissau, 2004). These developments led to an expansion of the stereotype threat literature, in which several moderator and mediator variables were studied.

Of all the studied moderator and mediator variables, we will summarize those variables that have been studied most frequently. Item difficulty appears to moderate the effects of stereotype threat, with difficult items leading to stronger effects (Campbell & Collaer, 2009; O'Brien & Crandall, 2003; Spencer et al., 1999; Wicherts, Dolan, & Hessen, 2005). Test-takers who are strongly identified with the relevant domain, in this case the domain of mathematics, science or spatial skills, appear to show stronger stereotype threat effects (Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Lesko & Corpus, 2006; Pronin, Steele, & Ross, 2004; Steinberg, Okun, & Aiken, 2012). Another theoretical moderator is gender identification; the effects of stereotype threat are generally more severe for women who are highly gender-identified (Kiefer & Sekaquaptewa, 2007; Rydell, McConnell, & Beilock, 2009; Schmader, 2002; Wout, Danso, Jackson, & Spencer, 2008). However, the latter results were contradicted in a Swedish study (Eriksson & Lindholm, 2007). Moreover, the effects of stereotype threat appear stronger within a threatening environment (e.g., in the presence of men, or when negatively stereotyped test-takers hold a minority status) compared to a safe environment (e.g., in the presence of women only, or when holding a majority status; Gneezy, Niederle, & Rustichini, 2003; Inzlicht, Aronson, Good, & McKay, 2006; Inzlicht & Ben-Zeev, 2003; Sekaquaptewa & Thompson, 2003). The presence of role models also appears to moderate the effect of stereotype threat, in such a way that role models that contradict the stereotype (i.e., women who are good in mathematics or men who lack mathematical skills) appear to protect females from the debilitating effects of stereotype threat on MSSS test performance (Elizaga & Markman, 2008; Marx & Ko, 2012; Marx & Roman, 2002; McIntyre, Paulson, Taylor, Morin, & Lord, 2011; Taylor, Lord, McIntyre, & Paulson, 2011). Finally, several researchers suggested that the stereotype threat effect is (partly) mediated by arousal (Ben-zeev, Fein, & Inzlicht, 2005), anxiety and worries (Brodish & Devine, 2009; Ford, Ferguson, Brooks, & Hagadone, 2004; Gerstenberg, Imhoff, & Schmitt, 2012; Osborne, 2001, 2007), or the occupation of working memory (Beilock, Rydell, & McConnell, 2007; Bonnot & Croizet, 2007; Rydell, Rydell, & Boucher, 2010; Schmader & Johns, 2003).

The literature on the effects of stereotype threat has been summarized by five meta-analyses that covered heterogeneous subsets of studies (Nguyen & Ryan, 2008; Picho et al., 2013; Stoet & Geary, 2012; Walton & Cohen, 2003; Walton & Spencer, 2009). These broad-stroke meta-analyses estimated a small to medium significant effect before moderators were taken into account, with standardized mean differences ranging from 0.24 (Picho et al., 2013) to 0.48 (Walton & Spencer, 2009). These findings seemed to confirm that the effect is rather stable, although most of these meta-analyses reported heterogeneity in effect sizes (Picho et al., 2013; Stoet & Geary, 2012; Walton & Cohen, 2003). In fact, the previous meta-analyses included diverse tests, settings, and stereotyped groups, which makes it hard to pinpoint exactly why some studies show larger effects than others. Although these large scale meta-analyses are interesting to portray an overall picture, a more homogeneous subset of studies is preferred when dealing with specific questions, like the degree to which the stereotype threat related to gender also influences MSSS performance in schools. Thus, we addressed this issue by selecting a specific stereotyped group and stereotype (i.e., women and their supposed inferior capacity of solving mathematical or spatial tasks) and a specific age group (i.e., those younger than 18 years), which should result in a less heterogeneous set of effect sizes. These design elements enabled us to describe the influence of stereotype threat on MSSS test performance for females in critical periods of human development, namely childhood and adolescence.

1.2. Stereotype threat and children

Although the effects of stereotype threat on women was traditionally studied within adult populations (Spencer et al., 1999), multiple studies over the last 15 years have been carried out with children and adolescents as participants (e.g., Ambady, Shih, Kim, & Pittinsky, 2001; *Keller & Dauenheimer, 2003). Studies on children and adolescents in schools contribute to the literature for at least three reasons: (1) to find out at which age the stereotype threat effect actually emerges, (2) to study the stereotype threat effect in the natural setting of the classroom instead of the laboratory setting, and (3) to address the question whether variables that moderate the stereotype threat effect in adult samples similarly moderate the stereotype threat effect among children.