

# The influence of stereotype threat on immigrants: review and meta-analysis

Markus Appel<sup>1\*</sup>, Silvana Weber<sup>1</sup> and Nicole Kronberger<sup>2</sup>

<sup>1</sup> Department of Psychology, University of Koblenz-Landau, Landau, Germany, <sup>2</sup> Department of Education and Psychology, Johannes Kepler University of Linz, Linz, Austria

## OPEN ACCESS

### Edited by:

Douglas Kauffman,  
Boston University School of Medicine,  
USA

### Reviewed by:

Angelica Moè,  
Università di Padova, Italy  
Claudio Longobardi,  
University of Turin, Italy

### \*Correspondence:

Markus Appel,  
Department of Psychology, University  
of Koblenz-Landau, Fortstrasse 7,  
76829 Landau, Germany  
appelm@uni-landau.de

### Specialty section:

This article was submitted to  
Educational Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 12 April 2015

**Accepted:** 15 June 2015

**Published:** 08 July 2015

### Citation:

Appel M, Weber S and Kronberger N  
(2015) The influence of stereotype  
threat on immigrants: review  
and meta-analysis.  
*Front. Psychol.* 6:900.  
doi: 10.3389/fpsyg.2015.00900

In many regions around the world students with certain immigrant backgrounds underachieve in educational settings. This paper provides a review and meta-analysis on one potential source of the immigrant achievement gap: stereotype threat, a situational predicament that may prevent students to perform up to their full abilities. A meta-analysis of 19 experiments suggests an overall mean effect size of 0.63 (random effects model) in support of stereotype threat theory. The results are complemented by moderator analyses with regard to circulation (published or unpublished research), cultural context (US versus Europe), age of immigrants, type of stereotype threat manipulation, dependent measures, and means for identification of immigrant status; evidence on the role of ethnic identity strength is reviewed. Theoretical and practical implications of the findings are discussed.

**Keywords:** stereotype threat, immigrants, meta-analysis, identity

## Introduction

In most nations, students with an immigration background score lower in achievement tests than non-immigrant students, and they leave school earlier (OECD, 2010). This status quo is particularly troublesome as the percentage of children and adolescents with an immigrant background is growing in many countries around the world, and immigration will likely be an even more prevalent phenomenon in the future (OECD, 2013). Consequently, the immigrant achievement gap is on the agenda of politicians, the general public, and social scientists alike.

Many immigrant groups are faced with negative achievement stereotypes in the countries they live in; these negative achievement stereotypes typically address those immigrant groups that indeed underperform. Our focus lies on the consequences of being exposed to a negative group stereotype regarding cognitive performance. According to stereotype and social identity threat theory and research, salient negative stereotypes can undermine the performance of negatively stereotyped group members due to an extra pressure not to fail (Steele and Aronson, 1995; Steele et al., 2002; Inzlicht and Schmader, 2012). Most of the experimental evidence on stereotype threat is based on women (stereotype: low ability in numerical domains) and African Americans (stereotype: low intellectual ability). Does stereotype threat affect the performance of immigrant students? In order to clarify the influence of stereotype threat on immigrants, the present article provides the first comprehensive review and meta-analytic summary of research (published and unpublished) on stereotype threat effects among this population.

## Immigration, Achievement, and Stereotype Threat

Stereotype threat is conceived as a state of psychological discomfort that is thought to arise when individuals are confronted with a negative stereotype about their own group in a situation in which the negative stereotype could be confirmed (Steele and Aronson, 1995; Steele et al., 2002). According to an integrative model of stereotype threat (Schmader et al., 2008; Schmader and Beilock, 2012) this state is characterized by the interplay of a physiological stress response, increased monitoring of the performance situation, and the regulation of negative thoughts and emotions. These processes consume working memory capacity, which is unavailable for the task at hand. The reduced working memory in turn leads to underperformance in cognitively challenging tasks (e.g., Schmader and Johns, 2003; Beilock et al., 2007). Stereotype threat is best known for its influence in testing situations. In line with the theoretical framework (e.g., Steele, 1997; Steele et al., 2002) several recent empirical studies furthermore connected stereotype threat to poorer learning and disidentification from school (e.g., Rydell et al., 2010; Appel et al., 2011; Taylor and Walton, 2011; Appel and Kronberger, 2012).

Stereotype threat theory and research has put little emphasis on the distinction of stereotyped groups. Indeed, many of the main findings have been demonstrated for samples of women as well as for samples of African Americans, despite important differences in stereotype content and breadth (Shapiro and Neuberg, 2007; Logel et al., 2012). Although Latinos in the US and immigrant groups in other countries outside the US are regularly mentioned in theoretical pieces and overview papers as potential targets of stereotype threat (cf. Inzlicht and Schmader, 2012), empirical studies are rare and an integrative review is missing. The passing attention paid to this group is particularly noteworthy in view of the overwhelming number of stereotype threat studies published in recent years. In stark contrast to this neglect are the systematic detrimental effects suffered by immigrants in educational systems around the globe. In most OECD countries, first generation immigrants (i.e., foreign-born immigrants) achieve lower scores in ability tests than students without immigration backgrounds. Second generation immigrants (i.e., immigrated before the age of six or at least one parent foreign-born) tend to score higher than first generation immigrants, but still fall short of non-immigrants (OECD, 2010). While language problems and low socio-economic status may be responsible for large parts of the achievement gap, a substantial part of the variance remains to be explained. Thus, it is of high interest to examine whether stereotype threat affects the cognitive performance of diverse immigrant groups in the same way as other minorities. If stereotype threat theory and research applied to immigrants, these stereotypes could prevent immigrants to perform up to their abilities (Steele and Aronson, 1995; Steele et al., 2002), they could inhibit them at times of preparation and learning (Rydell et al., 2010; Appel and Kronberger, 2012), and finally could contribute to a disidentification from school and academic achievements (Steele, 1992, 1997). Before presenting an overview on the current meta-analysis, in the following we discuss

some of the difficulties, challenges, and open questions for immigration research and the implications these entail for our meta-analysis.

## The (Special?) Case of Immigrants

An important difficulty for stereotype threat research when applied to the field of immigrants is that many, but not all immigrant groups are faced with negative achievement stereotypes. In the past five decades, the majority of immigrants to the US have originated from either Asia or Latin America (US Department of Homeland Security, 2013). However, while Hispanic Americans are likely to be met with negative stereotypes in academic contexts (e.g., Nadler and Clark, 2011), Asian immigrants often are met with even superior expectations (e.g., Shih et al., 1999; Cheryan and Bodenhausen, 2000; Shih et al., 2002). Similarly, Northern and Western Europe – once a region where numerous people migrated from (cf. Hatton and Williamson, 1994) – over the past half a century has become an important destination for immigration, attracting people from Southern Europe, Turkey, Northern Africa, and former overseas colonies (e.g., Institute National de la Statistique et des Etudes Economiques [INSEE], 2013). Again, the stereotypic expectations directed at the various groups differ. While Turks in Germany, for example, are seen as “not willing to adapt” and “underdeveloped,” Italians are regarded as “well educated” (Jäckle, 2008); in Spain, immigrants from Latin America are stereotyped as being “lazy,” while Chinese immigrants have the image of being “hard working” and “smart” (Enesco et al., 2005). The examples not only illustrate the differential valence of stereotypes directed at different groups of immigrants, but also highlight that the content of stereotypes varies (Lee and Fiske, 2006). While some stereotypes concern cognitive and intellectual ability, others address aspects such as the willingness to integrate or diligence.

Due to the heterogeneity in stereotype valence and content, it is difficult to draw a coherent picture of the influence of stereotype threat on immigrants in general. However, it seems safe to say that Latin Americans in the US and Spain (Enesco et al., 2005; Tenenbaum and Ruck, 2007), and immigrants from Turkey, the Maghreb region, and the Balkan in Northern and Western Europe are likely to be characterized as underachieving at school (Verkuyten and Kinket, 1999; Kahraman and Knoblich, 2000; Jäckle, 2008). Accordingly, standardized testing suggests that it is exactly these groups that are most likely to show impaired performance compared to non-immigrants or more positively stereotyped immigrant groups (OECD, 2010). These are the groups this meta-analysis focuses on. African Americans are not included in the current meta-analysis because their history in the US dates further back in time and is different from other more recent immigrant groups, due to 300 years of slavery and another 100 years of post-slavery exploitation. This group also has been addressed already in great detail by prior stereotype threat theory and research (for an empirical overview, see Nguyen and Ryan, 2008).

A further challenge for research is that immigration status is a more complex and fuzzy category than other more visible and stable stereotype-relevant characteristics, such as gender or

skin color. While it is extremely difficult for a man to ever count as a woman (or to be considered male and female at the same time), it is possible for many groups of immigrants to merge into the respective recipient culture (while potentially remaining identified with the culture of origin). This acculturation process may be easier for some groups of immigrants than for others, whereby the differential nature of stereotypes directed at them might be one reason for their difficulties. However, it is possible that due to the acculturation process at some point they are regarded, and regard themselves, as first and foremost belonging to the residence country and not any more as an immigrant<sup>1</sup>. This categorical malleability entails the question of how to define immigration status in empirical research; it has been suggested that being an immigrant is not only an objective characterization, but also, and even more importantly so, a subjective state of ethnic identification (Deaux, 1996).

Furthermore, and in an attempt to acknowledge the complex nature of defining immigrants, research has described the psychological consequences of migrating from one culture to the other, or being born as a member of an immigrant group along the lines of two identity dimensions (e.g., Berry, 1997; Phinney et al., 2001). One dimension is the attachment to one's ethnic background of provenance, which includes the exploration of cultural practices of the culture of origin and the commitment to this cultural group. The second dimension is the attachment to one's culture of residence. Both dimensions are considered to be conceptually independent (Berry, 1990, 1997; Liebkind, 2001; Phinney et al., 2001). Subject to the identity strength on both dimensions, four different "acculturation profiles" have been described: *assimilation* (low ethnic origin identity, high residence culture identity), *separation* (high ethnic origin identity, low residence culture identity), *marginalization* (low ethnic origin identity, low residence culture identity), and *integration* (high ethnic origin identity, high residence culture identity). *Bicultural identity integration* (Benet-Martinez and Haritatos, 2005; Nguyen and Benet-Martinez, 2010) is an alternative conceptualization, in which different ways to deal with the two potential group identities (ethnic origin and residential background) are distinguished: for some individuals, both identities may be perceived as compatible and overlapping, for others, both identities can be perceived as non-overlapping and a potential source of conflict.

Group identification is important because of its psychological consequences. Immigration research indicates, for example, that ethnic identification can buffer the negative effects elicited by societal devaluation and rejection of immigrants (e.g., Romero and Roberts, 2003; Umaña-Taylor and Updegraff, 2007; Armenta and Hunt, 2009; Umaña-Taylor et al., 2012). Similarly, ethnic identification is likely to play an important role when it comes to stereotype threat. Stereotype threat has been conceived as a result of a cognitive imbalance between the concept of

self, the concept of a group, and the concept of an ability domain (Schmader et al., 2008). Stereotype threat occurs when individuals identify with a group (positive association between self and group) and identify with an ability domain (positive association between self and ability domain), while a negative stereotype suggests a negative connection between one's group and the domain at hand. Prior research on stereotype threat has frequently addressed the moderating role of domain identification, that is, the self-domain link (for a review see Nguyen and Ryan, 2008). The negative effects of stereotype threat are more pronounced when individuals identify positively with the domain in question, either because it is part of their chronic self-concept (e.g., Aronson et al., 1999) or due to a situational prime of ego-involvement (Spencer et al., 1999)<sup>2</sup>. The self-group link has received much less attention, although the stereotype threat process model (Schmader et al., 2008; Schmader and Beilock, 2012) posits that stereotype threat effects are particularly strong when the connection to one's group is highlighted by a situational prime; stereotype threat effects are also supposed to grow with the extent to which individuals identify with their stereotyped group. The empirical results on the self-group link are somewhat mixed. The threat increasing function of group identification has been demonstrated for women (Schmader, 2002) and African Americans (Ho and Sidanius, 2010), but other findings suggest that a strong identity might buffer stereotype threat effects among both groups (McFarland et al., 2003; Eriksson and Lindholm, 2007).

Extending the scope of stereotype threat theory to immigrants leads to the necessity of exploring the (potentially special) conditions and underlying mechanisms that apply for this particular target group. Overall it seems that definitions of group membership and processes of group identification may be more complex and less clear cut for immigrants than for other groups addressed by stereotype threat research. As a consequence, open questions abound on how and when ethnic identification affects intellectual performance. Moreover, there is a need to inspect the role of identification with the culture of residence (or a compatible and overlapping bicultural identity, Benet-Martinez and Haritatos, 2005).

## Rationale and Overview

Stereotype threat is a widely studied psychological phenomenon with potentially large implications for educational practice and related policies (cf. Cohen et al., 2012). Immigrants are frequently claimed to be one of the target groups of stereotype threat, but a systematic overview of available studies is missing. Despite the existence of a number of meta-analyses on stereotype threat, as yet, no meta-analysis has explicitly addressed the influence of stereotype threat on immigrants. Two meta-analyses were presented by Walton and Cohen (2003), providing results on the stereotype lift phenomenon, and on the relationship between ability and performance among various stereotyped

<sup>1</sup>Note that a longer duration of time spent in a country does not necessarily need to make the situation better for those stigmatized immigrants. Second generation Afro-Caribbeans, for example, were found to be *more* rather than less susceptible to stereotype threat compared to first generation immigrants, because stereotypes toward African Americans applied (Deaux et al., 2007; see also Waters, 1999).

<sup>2</sup>Stereotype threat further depends on the extent individuals are generally aware of the negative stereotype (e.g., stigma consciousness, Brown and Pinel, 2003) and on the situational cues that activate or deactivate negative stereotypes (e.g., Steele and Aronson, 1995; Spencer et al., 1999).

and non-stereotyped groups (Walton and Spencer, 2009). Their emphasis, however, had not been on a comparison between differences among individuals under varied stereotype threat conditions, and no meta-analytic results were provided for immigrants. Nguyen and Ryan (2008) focused on stereotype threat effects, but no differentiation between African American and Latino American participants was made. Finally, Nadler and Clark (2011) examined stereotype threat effects among African Americans and Latinos in the US, but their sample relating to the latter group involved only four applicable studies (see Results) and stereotype threat research with immigrants in countries outside the US was not examined.

Beyond a quantification of the average effect size of stereotype threat research dealing with immigrants, we were interested in potential moderation effects. Heterogeneity in effect size often not only results from theoretically important variation, but also from characteristics of the samples included and from research practices (and potential related biases). To address the latter aspect, our first moderator was the publication status of the study. In a recent overview on stereotype threat research dealing with the math performance of female children and adolescents (Ganley et al., 2013), a null effect for unpublished research was reported. Thus, an analysis of moderation effects of publication status could provide a hint at potential publication bias.

We further examined whether results differed between studies with Latinos in the US versus different immigrant groups in European countries. Although stereotype threat generally got demonstrated in many countries, so far the large majority of studies has been conducted in the US. Cultural differences and variations in stereotype content might yield important differences in the threat experience and the effects on performance (cf. Eriksson and Lindholm, 2007). Thus, it was one of our main goals to identify whether or not stereotype threat effects on immigrants differed between cultures.

Research on stereotype threat is often based on undergraduate volunteers. We analyzed whether the study samples consisted of undergraduates or other samples like children and adolescents, because the experience of undergraduates might not be readily transferable to other samples (cf. Ganley et al., 2013).

We additionally analyzed whether different stereotype threat-activating cues, as defined by Nguyen and Ryan (2008), were used. In line with these authors, the following three categories were distinguished: *blatant* (explicit statement about the inferiority of one group, e.g., “women score lower in math than men”), *moderately explicit* (statement about subgroup differences in performance, but the direction of the differences is left open, e.g., “this test has shown gender differences in the past”), and *indirect and subtle* (i.e., no statement about subgroup differences; instead, the context of tests, test takers’ subgroup membership, or test taking experience is manipulated, e.g., test is “diagnostic” versus “not diagnostic”).

We were further interested in the tasks to assess performance. Are the results obtained equivalent for non-verbal tasks that assess basic cognitive abilities on the one hand (such as Raven’s progressive matrices) or tests that include more knowledge-based tasks on the other hand (such as GRE-typed measures)?

Finally, we closely inspected available studies for their way to assess immigrant group status. One method consists of participants self-identifying themselves as members of a certain immigrant group. However, another frequent option is to identify immigrants via demographic characteristics like the place of birth of themselves, their parents, or their grandparents. By employing the latter method, rather than self-categorization, individuals could be ascribed an immigrant background status, although they would self-define to belong to the mainstream culture. These individuals might be unaffected by a stereotype threat manipulation that targets their ethnic group of origin, due to the weak self-group link. As a consequence, studies that used methods other than self-identification might yield weaker effects than studies that relied on self-identification. Beyond immigrant status categorization, we were particularly interested in identity aspects, as previous research suggests competing predictions on the role of ethnic identity strength.

## Materials and Methods

### Literature Search and Selection Criteria

To identify relevant studies a literature search was conducted in December 2012, which was repeated in December 2013 and in March 2015. We searched the databases *PsycInfo*, *ERIC*, *SocIndex*, *PsynDEX*, and *Psychology and Behavioral Sciences Collection* for studies that contained the terms “stereotype threat” or “social identity threat” and at least one of the terms *migra\**, *immigr\**, *Latin\**, *Hispanic\**, or *Turk\** (asterisk as a placeholder for different word endings)<sup>3</sup>. Our database search for literature published until 2014 resulted in 120 references. In addition, *google scholar* was searched for documents that contained a combination of the terms “stereotype threat,” “immigrant,” or “migrant,” and “experiment.” We further inspected the reference lists of book chapters and review articles for further studies. Finally, we asked for additional published or unpublished studies through e-mailing lists.

Full texts for potentially relevant studies were retrieved. We included studies that met the following criteria (inclusion/exclusion criteria): first, participants or a distinct subgroup of participants were members of a group with a recent and ongoing immigration history, such as Latinos in the US or Arab immigrants in Europe (because the history of African Americans is much older and intertwined with the complexities of race and skin color, this group was not included). Thus, studies that manipulated the salience of a stereotype about immigrants, but focused exclusively on non-immigrants did not meet this criterion (e.g., Chatard et al., 2008). Likewise, studies that did not report separate results for immigrants and other groups, most notably studies that reported combined average scores for Latino and African American participants were not included (e.g., Howard and Anderson, 2010). Second, the activated immigrant stereotype needed to be negative. This excluded research on the consequences of stereotypes regarding

<sup>3</sup>We further made sure that Arab\* and African\* immigrants to Europe were covered by the search procedure; only the group of African Americans was to be excluded. No additional studies were found.

Asian immigrants (e.g., Cheung and Schmader, unpublished data), as widely held stereotypes regarding the cognitive performance of Asians are rather positive (e.g., Cheryan and Bodenhausen, 2000; Shih et al., 2002). Third, the study had to be experimental and had to follow the stereotype threat or social identity threat family of experimental treatments. Immigrant participants were supposed to be randomly assigned to two or more experimental groups. The conditions differed with respect to the salience of a negative stereotype addressing the immigrant group, the supposed relevance of the negative stereotype for an upcoming task (relevance), or the presence/absence of stimuli that signal non-belonging or belonging to a domain or society at large. Fourth, a measure of cognitive performance served as a dependent variable. Fifth, we inspected all studies for the quality of the applied methods and measures. This included an analysis of the operationalization of the independent variable and the dependent measures. Specifically, theory-based preconditions for stereotype threat to occur were inspected, such as sufficient domain identification and substantial task difficulty (Schmader et al., 2008).

We identified 18 texts (published and unpublished articles and dissertations) containing 21 experiments that met our criteria. In addition to 19 English-language texts, one report was written in German, and one in French (both languages were intelligible to us). In two cases the identical study was used for two separate journal articles. These results were included in our analysis only once. Thus, our final sample consisted of 19 experiments, 10 of which were unpublished (Table 1).

### Effect Size Coding and Procedure

All three authors read and coded all available studies. A coding sheet was developed to gather the relevant information. Discrepancies between the judgments were very rare and resolved through discussions. Due to the fact that the examined datasets involved a comparison between two groups – stereotype threat high or low – the standardized mean difference was chosen as the effect size measure (Cohen's *d*). When available, the effect size calculations were based on descriptive data (*Ms*, *SDs*, *ns*), when unavailable, formulas to calculate the standardized mean difference based on *t*-test statistics or *F*-statistics with 1 degrees of freedom were employed (Lipsey and Wilson, 2001; Wilson, 2012). We ensured that our standardized mean difference score reflected the mean difference between immigrants under conditions of low versus high stereotype threat. Our standardized mean difference scores never represented an interaction effect (e.g., stereotype threat treatment by immigrant background). We did not consider such interaction effects because they can be driven – in part or completely – by stereotype lift effects (Walton and Cohen, 2003) among non-immigrant groups. We also did not consider multi-group comparisons; when the stereotype threat treatment involved more than one group (e.g., reflecting stereotype threat low, intermediate, and high) those two groups were compared that should theoretically exhibit maximum difference. In some studies several performance scores were reported (e.g., scores for two or more subtests of an intelligence test, Pellegrini, 2005; Wicherts et al., 2005). In order to preserve independence of effect sizes, we averaged the

scores before they were included in the meta-analysis (Headrick, 2010).

## Coding of Study Characteristics

### Sample Background

We distinguished two broader groups of studies based on their origin and sample: those that addressed stereotype threat among Latinos in the US ( $k = 11$ ) and studies that focused on immigrants in Europe ( $k = 8$ ). The latter studies were conducted in Austria, Belgium, France, Germany, or the Netherlands.

### Circulation

Whereas nine studies were published in outlets of educational or social psychology, a majority of 10 studies was unpublished.

### Age Group

All studies conducted in the US and four of the European studies investigated samples of young adults, typically consisting of undergraduates ( $k = 14$ ). Five of the European studies investigated adolescents.

### Stereotype Threat Treatment

Based on the distinction of stereotype threat-activating cues in the meta-analysis by Nguyen and Ryan (2008), out of the 19 experiments included in our meta-analysis, seven studies were found to have used indirect and subtle cues, five used moderately explicit cues, and seven used blatant stereotype threat-activating cues.

### Dependent Measures

We distinguished studies that used measures of knowledge and abilities such as GRE-type tasks (verbal or maths,  $k = 13$ ) from studies that used non-verbal measures of general cognitive functioning, such as (working) memory tests, or tests in the tradition of Raven's progressive matrices ( $k = 6$ ).

### Group Categorization

There are different ways to distinguish immigrant group members from non-immigrants. Demographic surveys and statistical reports (e.g., OECD, 2010) typically rely on the place of birth of the individuals or their parents. Another common method is to ask individuals with which group they identify most. Studies were coded whether or not self-identification was the means to distinguish between immigrant group members or non-members (self-identification,  $k = 13$ ).

### Comparison with Previous Meta-Analyses

Our dataset diverges remarkably from the data obtained in previous meta-analyses on stereotype threat. In addition to differences in the main aims and methodological approaches outlined above (focus on immigrants, quantification of stereotype threat main effects, and moderating variables), there are specific differences that appear noteworthy. We decided to meta-analyze the main effects or the simple main effects of the stereotype threat treatment on immigrants. As a consequence, the effect sizes integrated in our meta-analysis differ in part from the effect sizes reported in the meta-analysis by Nguyen and Ryan (2008). One previous meta-analysis (Nadler and Clark, 2011)

**TABLE 1 | Characteristics of the studies included in the meta-analysis.**

| Study no. | Study  | P   | Immigrant group                                   | Manipulation of stereotype or social identity threat   | Dependent measure  | DV group |
|-----------|--|-----|---|--|--|----------|
| 1         | Appel (2012)                                     | Yes | Immigrants of different origin in Austria         | Anti-immigrant or neutral political party posters were shown in-between two intelligence tests   | CFT  | 0        |
| 2         | Armenta (2010)                                   | Yes | Latino/as in the US                               | Test characterized either as diagnostic of intelligence and sensitive to ethnic differences or test purpose stated to be computerized versus paper-pencil  | Math test (algebra, geometry, similar to GRE)                                      | 1        |
| 3         | Berjot et al. (2011)                             | Yes | North Africans in France                          | Test characterized either as diagnostic of intelligence or as a memory test. Ethnicity asked before or after test.   | Memory test (Rey-figure)   | 0        |
| 4         | Chateignier et al. (2009)                        | Yes | Arabs in France                                   | Test characterized either as diagnostic of intelligence or as characterized as a memory test in development  | Verbal ability items (GRE)   | 1        |
| 5         | Flanagan (2009)                                  | No  | Latino/as in the US                               | Classroom notes were shown in order to remind participants of ethnic inequalities or neutral condition   | Verbal ability items (similar to GRE)  | 1        |
| 6         | Froehlich et al. (in revision)                   | No  | Turks in Germany                                  | Test characterized either as diagnostic of verbal intelligence or as test under development  | Verbal ability items (similar to GRE)  | 1        |
| 7         | Gonzales et al. (2002)                           | Yes | Latino/as in the US                               | Test characterized either as diagnostic or as non-diagnostic of mathematical and spatial abilities   | Mathematical and spatial ability (Wonderlic test)                                  | 1        |
| 8         | Hollis-Sawyer and Sawyer (2008)                  | Yes | Latino/as in the US                               | Test characterized either as diagnostic of 'general ability' or as an interest measure   | CFT  | 0        |
| 9         | Klein et al. (2007)                              | Yes | Sub-Saharan Africans in Belgium                   | Test characterized as an entrance test for prestigious jobs. It was or was not further stated that Africans were found to underperform in this test  | CFT  | 0        |
| 10        | Mok et al. (in preparation)                      | No  | Turks in Germany                                  | Introduction stated that test has (not) revealed performance differences between Germans and Turks in the past   | Verbal ability items (GRE)   | 1        |
| 11        | Pellegrini (2005)                                | No  | Latinas in the US                                 | Test characterized either as diagnostic or as non-diagnostic of "true abilities"   | WAIS (subtests letter-no., similarity, block, arithmetic)                          | 1        |
| 12        | Salinas (1998), Study 1                          | No  | Latino/as in the US                               | Participants were or were not asked to rate the "level of bias" of the test prior to the crucial test part. Moreover, participants were or were not applied to electrodes and connected to a pseudo "effort meter" | Verbal ability items (GRE, computerized)   | 1        |
| 13        | Salinas (1998), Study 2                          | No  | Latino/as in the US                               | Participants were or were not asked to rate the "level of bias" of the test, prior to the crucial test part  | Verbal ability items (GRE)   | 1        |
| 14        | Schmader and Johns (2003), Study 2               | Yes | Latino/as in the US                               | Task characterized either as predictive of intelligence and meant to establish group-wise norms or no such information   | Working memory test  | 0        |
| 15        | Schultz et al. (unpublished manuscript), Study 1 | No  | Latino/as in the US                               | Introduction mentioned previous underperformance of Latinos and participants had to list three reasons for that. In the control condition neither was this information given nor the task required                 | Verbal ability items (similar to GRE)  | 1        |
| 16        | Schultz et al. (unpublished manuscript) Study 2  | No  | Latino/as in the US                               | Same as in Schultz et al. (unpublished manuscript), Study 1  | Verbal ability items (similar to GRE)  | 1        |
| 17        | Schultz et al. (unpublished manuscript) Study 3  | No  | Latino/as in the US                               | Same as in Schultz et al. (unpublished manuscript), Study 1  | Verbal ability items (similar to GRE)  | 1        |
| 18        | Stünzendorfer (2006)                             | No  | Turks in Germany                                  | Test characterized either as an ability test that shows differences between immigrants and non-immigrants or as a test non-diagnostic of abilities   | Standard Progressive Matrices (Raven, 1958)  | 0        |
| 19        | Wicherts et al. (2005)                           | Yes | Immigrants of different origin in the Netherlands | Test characterized either as an intelligence test and ethnicity items placed before the DV or test not characterized as an intelligence test and ethnicity items placed after DV                                   | Dutch intelligence test (adapted), numerical, abstract reasoning, verbal reasoning | 1        |

P, Published?; CFT, "Culture Fair Test" (based on Cattell, 1940); WAIS, Wechsler Adult Intelligence Scale; DV, Dependent variable: 0, non-verbal fluid intelligence or memory tasks, 1, GRE-like measures (verbal, maths).

identified six studies on stereotype threat effects among Latino samples in the US. We did not include two of these six studies in our analysis, because they did not meet our inclusion criteria. In one study (Stone, 2002) a negative stereotype with

respect to White Americans was examined, Latinos served as a control group. In the second study (Good et al., 2003), treatment effects among Latino, Black, or White students were not distinguished. Thus, only four out of the 19 identified studies were already included in the meta-analysis by Nadler and Clark (2011).

## Results

### Meta-Analytic Results: Average Effects Effect Size of the Stereotype Threat Treatment

The meta-analytic procedure followed the recommendations by Lipsey and Wilson (2001; Wilson, 2012). All standardized effect sizes were adjusted for small sample bias (Hedges, 1981). The inverse variance served as a weight that was allotted to each effect size (see Lipsey and Wilson, 2001, for the respective formulas). Negative effect sizes indicate a worse performance in the stereotype threat high than in the stereotype threat low condition (Table 2). All studies except for one (Wicherts et al., 2005, Study 1) had a negative effect size, indicating that in the great majority of studies the descriptive means followed the pattern expected from stereotype threat theory. The unweighted mean of the studies' effect sizes amounted to  $M = -0.68$

(SD = 0.54). With respect to outliers, the individual effect size of one study (Berjot et al., 2011) was around  $M \pm 3$  SD of the average unweighted mean effect size. No other study surpassed the threshold of  $M \pm 2$  SD. Results with and without this particular study were inspected.

The weighted average effect of the stereotype threat manipulation over 19 independent effect sizes amounted to  $d = -0.63$  for the random effects model (95% CI = -0.86; -0.40), SE = 0.12,  $Z = -5.36$ ,  $p < 0.001$ , and  $d = -0.49$  (95% CI = -0.61; -0.36) for the fixed effects model, SE = 0.06,  $Z = -7.68$ ,  $p < 0.001$ . This indicates an average effect in support of stereotype threat theory among immigrant samples. This effect holds if the outlier study is excluded, random effects model;  $d = -0.52$  (95% CI = -0.71; -0.34), SE = 0.09,  $Z = -5.58$ ,  $p < 0.001$ , fixed effects model:  $d = -0.44$  (95% CI = -0.56; -0.31), SE = 0.06,  $Z = -6.80$ ,  $p < 0.001$ . According to the interpretation by Cohen (1988), the overall effect is medium in size. We further examined the homogeneity of our sample of effect sizes. With  $Q(18) = 57.04$ ,  $p < 0.001$ , the effect sizes were significantly heterogeneous, highlighting the importance of subsequent moderator analyses.

### Analysis of Sampling Bias and File-Drawer Analysis

Our goal in the meta-analysis was to include a maximum of studies, published or unpublished, written in English, Spanish, German, or French. More than half of our datasets originated from unpublished research. As reported in the previous section, these studies' average effect is significant and moderate in size. Despite our efforts, however, it is unlikely that we were able to uncover every study conducted so far that would have met our criteria.

In order to estimate a potential sampling bias in our set of studies we first plotted the meta-analytic data for a visual inspection (Figure 1). The funnel plot illustrates that the great

TABLE 2 | Study results and statistics.

| Study no.         | Immigrant sample size | SMD (d) | SE SMD (d) | $w_{iv}$ |
|-------------------|-----------------------|---------|------------|----------|
| 1 <sup>a</sup>    | 49                    | -0.41   | 0.29       | 11.96    |
| 2 <sup>b</sup>    | 40                    | -1.45   | 0.36       | 7.92     |
| 3                 | 44                    | -2.37   | 0.39       | 6.46     |
| 4 <sup>c</sup>    | 50                    | -0.72   | 0.29       | 11.75    |
| 5                 | 36                    | -0.47   | 0.34       | 8.75     |
| 6                 | 126                   | -0.02   | 0.18       | 31.47    |
| 7                 | 60                    | -0.70   | 0.27       | 14.13    |
| 8                 | 47                    | -1.13   | 0.31       | 10.11    |
| 9 <sup>d</sup>    | 34                    | -0.64   | 0.35       | 8.09     |
| 10                | 78                    | -0.21   | 0.23       | 19.26    |
| 11                | 60                    | -0.51   | 0.26       | 14.53    |
| 12 <sup>e,f</sup> | 20                    | -0.92   | 0.47       | 4.48     |
| 13 <sup>e</sup>   | 56                    | -0.67   | 0.28       | 13.18    |
| 14                | 33                    | -0.54   | 0.35       | 7.96     |
| 15                | 44                    | -0.68   | 0.31       | 10.39    |
| 16                | 40                    | -0.64   | 0.32       | 9.51     |
| 17                | 80                    | -0.33   | 0.23       | 19.72    |
| 18 <sup>g</sup>   | 25                    | -0.63   | 0.42       | 5.73     |
| 19                | 138                   | 0.06    | 0.17       | 34.37    |

SMD, standardized mean difference (after adjustment for small sample size bias, Hedges, 1981);  $w_{iv}$ , inverse variance weight.

<sup>a</sup>Effect size was derived from change score with additional information not included in the article:  $r_{pre/post}$  (control condition) = 0.85;  $r_{pre/post}$  (stereotype threat condition) = 0.66.

<sup>b</sup>Descriptives for men and women were averaged.

<sup>c</sup>Equal group size was presumed.

<sup>d</sup>"Control group" as low stereotype threat group.

<sup>e</sup>Only post-test scores were analyzed.

<sup>f</sup>ST+EM (effort meter) served as maximum ST-condition.

<sup>g</sup>Based on t-test results ( $p = .83$ ) and reduced sample size.

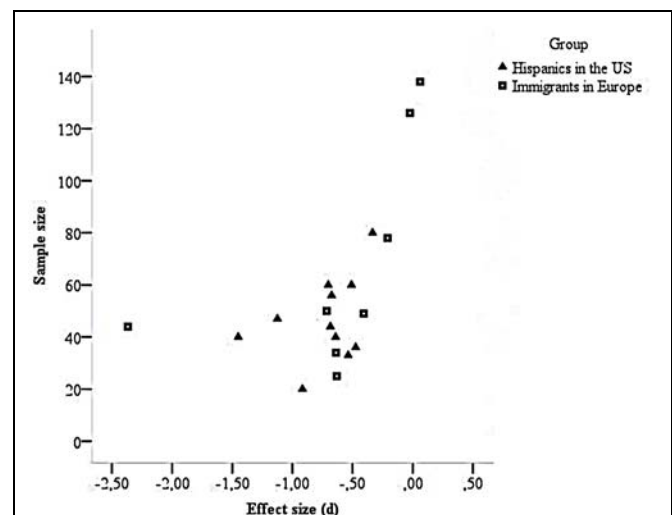


FIGURE 1 | Funnel plot based on effect size (d) and sample size. In studies with negative effect sizes, low stereotype threat groups outperformed high stereotype threat groups.

majority of studies yielded a negative effect size, indicating that the direction of the effect was regularly in support of stereotype threat effects. It further shows that studies with larger samples were more likely to yield null effects. The plot points at a lack of small-sample studies with effects that do not support a stereotype threat hypothesis. One reason for this finding could be a selective reporting of small-scale studies. In order to gage the file drawer problem (Rosenthal, 1979), we first calculated the number of studies confirming the null hypothesis that would be needed to conclude that the effect is small. We used a formula by Orwin (1983; Lipsey and Wilson, 2001) and we set the limit of a small effect size at  $d = 0.2$  (Cohen, 1988). This analysis yielded a fail-safe number of 46 unaccounted for studies in support of the null hypothesis that were needed to reduce the effect size to  $d = 0.2$ . When the limit was set at an arguably insubstantial average effect size of  $d = 0.1$ , a fail-safe number of 111 unaccounted for studies in support of the null hypothesis was identified.

Taken together, our sampling analysis pointed out a remarkable lack of null effects in small sample studies. If such studies were conducted, they were unavailable to us. A file-drawer analysis showed that the number of studies in support of the null hypothesis that were needed to change the average effect size to small or even to insubstantial is rather large. Thus, we conclude that the average effect size in support of a stereotype threat effect among people with an immigrant background is not severely challenged by potentially existing but unaccounted for studies. As a complement to our discussion on publication bias, the following moderator analyses included publication status (published versus unpublished) as one possibly influential factor.

### Meta-Analytic Results: Moderator Analyses

Due to the fact that the effect sizes were significantly heterogeneous, we inspected the influence of factors with theoretical relevance. This included the focused immigrant group (Latinos in the US versus immigrants in Europe), participants' age, the experimental treatment, the dependent variable, and method of group categorization (see **Table 3**). Before turning to these conceptually relevant aspects, differences between published and unpublished studies are addressed.

#### Publication

More than half of our studies were unpublished, so it seemed warranted to contrast published with non-published effects. The moderator analysis yielded a non-significant difference,  $Q(1) = 2.00$ ,  $p = 0.157$ , indicating that the effect sizes between both groups did not significantly differ. This result holds with or without the study by Berjot et al. (2011). The average standardized mean differences were significant both for published studies,  $d = -0.81$ ,  $SE = 0.17$ ,  $p < 0.001$ , as well as for unpublished research  $d = -0.47$ ,  $SE = 0.16$ ,  $p = 0.004$ .

#### Latinos in the US versus Immigrants in Europe

One main goal of the meta-analysis was to examine potential differences in effect size between studies conducted with Latino samples in the US and immigrants in Europe in order to gage whether stereotype threat is a sufficiently replicated phenomenon with immigrant samples outside the US.

Based on the 19 studies, there were no significant differences between the two groups,  $Q(1) = 0.79$ ,  $p = 0.374$ . Separate calculations showed that the average effects were significant for Latinos in the US,  $d = -0.71$ ,  $SE = 0.15$ ,  $p < 0.001$ , as well as for immigrants in Europe  $d = -0.50$ ,  $SE = 0.17$ ,  $p = 0.003$ , indicating equivalent support for stereotype threat theory. Studies with Latinos in the US showed homogeneous effect sizes,  $Q(10) = 4.06$ ,  $p = 0.945$ . For studies with immigrants in Europe, a significant score,  $Q(7) = 15.37$ ,  $p = 0.032$ , pointed at remarkable heterogeneity. In a subsequent step, we tried to examine whether the heterogeneity could be attributed to the one study with an exceptionally large effect size (Berjot et al., 2011) which was situated in Europe. Would the significance of average mean difference scores for European studies still hold, if this study was eliminated from the dataset? The re-analysis yielded a significant moderation effect of immigrant group,  $Q(1) = 9.12$ ,  $p = 0.003$ , indicating that the effect sizes for studies conducted in Europe were lower than the effect sizes for US studies. When the extreme score of the study by Berjot et al. (2011) was excluded, the average standardized mean difference dropped for the European study sample, but was still significant,  $d = -0.24$ ,  $SE = 0.11$ ,  $p = 0.026$ . This additional analysis pointed to consistent stereotype threat effects for both US Latino and European immigrant samples, even if the effects obtained from the latter group might be weaker than those obtained from Latinos in the US.

All results for the following moderator analyses hold with or without the study by Berjot et al. (2011), so they are based on the full sample.

#### Age Group

As noted in the method section, the five datasets on adolescent samples under threat came from European studies. The moderator analysis revealed a larger effect among adults (which were mostly undergraduates) than among adolescents,  $Q(1) = 9.97$ ,  $p = 0.002$ . The present data of five independent studies allow no clear conclusion on whether stereotype threat affects immigrant adolescents or not,  $d = -0.17$ ,  $p = 0.294$ .

#### Treatment Type

In line with the findings from a previous meta-analysis (Nguyen and Ryan, 2008), moderately blatant stereotype threat treatments yielded the largest results [the difference between the three treatment-groups was trend-significant,  $Q(2) = 5.98$ ,  $p = 0.050$ ], but all types of treatment were effective (see **Table 3**).

#### Dependent Measures

We identified a trend-significant difference between the dependent measures,  $Q(1) = 2.94$ ,  $p = 0.086$ , suggesting that studies which used non-verbal, fluid intelligence measures (memory tasks, RPM-like tasks) obtained somewhat stronger effects than studies that used GRE-like tasks. The average influence of the stereotype threat treatment was significant for both groups of dependent variables.

#### Group Categorization

In order to specify immigrant status, thirteen studies relied on self-identification while six studies used other procedures,



**TABLE 3 | Summary of the moderator analyses, mixed effects model.**

| Variable                                       | <i>k</i> | Between-groups analysis       | Subgroup effect size  | By group analysis               |
|--|----------|-------------------------------|---|---------------------------------|
| Circulation                                    |          | Q(1) = 2.00, <i>p</i> = 0.157 |   |                                 |
| Unpublished                                    | 10       |                               | <i>d</i> = -0.47 (95% CI = -0.79; -0.15, SE = 0.16, <i>Z</i> = -2.91, <i>p</i> = 0.004) | Q(9) = 2.39, <i>p</i> = 0.984   |
| Published                                      | 9        |                               | <i>d</i> = -0.81 (95% CI = -1.14; -0.47, SE = 0.17, <i>Z</i> = -4.67, <i>p</i> < 0.001) | Q(8) = 14.01, <i>p</i> = 0.082  |
| Immigrant group                                |          | Q(1) = 0.79, <i>p</i> = 0.374 |   |                                 |
| Latina/o in the US                             | 11       |                               | <i>d</i> = -0.71 (95% CI = -1.00; -0.42, SE = 0.15, <i>Z</i> = -4.74, <i>p</i> < 0.001) | Q(10) = 4.06, <i>p</i> = 0.945  |
| Immigrants in Europe                           | 8        |                               | <i>d</i> = -0.51 (95% CI = -0.84; -0.18, SE = 0.17, <i>Z</i> = -3.01, <i>p</i> = 0.003) | Q(7) = 15.37, <i>p</i> = 0.032  |
| Age group                                      |          | Q(1) = 9.97, <i>p</i> = 0.002 |   |                                 |
| Adolescents                                    | 5        |                               | <i>d</i> = -0.17 (95% CI = -0.48; 0.15, SE = 0.16, <i>Z</i> = -1.05, <i>p</i> = 0.294)  | Q(4) = 1.99, <i>p</i> = 0.738   |
| Adults   | 14       |                               | <i>d</i> = -0.79 (95% CI = -1.01; -0.57, SE = 0.11, <i>Z</i> = -7.12, <i>p</i> < 0.001) | Q(13) = 17.57, <i>p</i> = 0.175 |
| Stereotype threat treatment                    |          | Q(2) = 5.98, <i>p</i> = 0.050 |   |                                 |
| Subtle   | 7        |                               | <i>d</i> = -0.62 (95% CI = -0.96; -0.27, SE = 0.18, <i>Z</i> = -3.51, <i>p</i> < 0.001) | Q(6) = 3.81, <i>p</i> = 0.702   |
| Moderate                                       | 5        |                               | <i>d</i> = -1.05 (95% CI = -1.49; -0.62, SE = 0.22, <i>Z</i> = -4.70, <i>p</i> < 0.001) | Q(4) = 9.95, <i>p</i> = 0.041   |
| Blatant  | 7        |                               | <i>d</i> = -0.36 (95% CI = -0.70; -0.04, SE = 0.17, <i>Z</i> = -2.19, <i>p</i> = 0.029) | Q(6) = 2.41, <i>p</i> = 0.879   |
| Dependent measures                             |          | Q(1) = 2.94, <i>p</i> = 0.086 |   |                                 |
| Non-verbal, fluid intelligence tasks or memory | 6        |                               | <i>d</i> = -0.93 (95% CI = -1.37; -0.48, SE = 0.23, <i>Z</i> = -4.05, <i>p</i> < 0.001) | Q(5) = 9.75, <i>p</i> = 0.083   |
| GRE-like (verbal, maths)                       | 13       |                               | <i>d</i> = -0.50 (95% CI = -0.75; -0.25, SE = 0.13, <i>Z</i> = -3.92, <i>p</i> < 0.001) | Q(12) = 8.59, <i>p</i> = 0.737  |
| Group categorization                           |          | Q(1) = 5.63, <i>p</i> = 0.018 |   |                                 |
| Immigrant status not self-identified only      | 6        |                               | <i>d</i> = -0.28 (95% CI = -0.61; 0.05, SE = 0.17, <i>Z</i> = -1.69, <i>p</i> = 0.090)  | Q(5) = 3.48, <i>p</i> = 0.626   |
| Immigrant status self-identified               | 13       |                               | <i>d</i> = -0.78 (95% CI = -1.02; -0.54, SE = 0.12, <i>Z</i> = -6.29, <i>p</i> < 0.001) | Q(12) = 8.59, <i>p</i> = 0.192  |

including, for example, the parents' birthplace. In the latter studies, some participants might be categorized into the immigrant group whereas they would have self-identified to be a majority group member, resulting in the violation of a relevant precondition for stereotype threat to occur, namely the identification with the stereotyped target group. Studies in which immigrant group members self-identified to belong to this group yielded larger effect sizes than studies in which self-identification was no prerequisite,  $Q(1) = 5.63, p = 0.018$ . This is in line with the assumption that a stereotype threat treatment might fail for individuals who technically fall into the immigrant group category, but self-identify to be a majority group member.

### Stereotype Threat and Identity Strength

Some research conducted outside the stereotype threat framework identified the strength of an immigrant's psychological ties to his or her ethnic group to be an adaptive factor that attenuates the influence of ethnicity-related stressors (e.g., Umaña-Taylor and Updegraff, 2007; Armenta and Hunt, 2009). Within the stereotype threat framework, however, identity strength is considered to increase threat and to reduce cognitive performance (Schmader et al., 2008).

Unfortunately, few studies in our sample addressed identity strength as a potential moderating factor. In the experiment by Armenta (2010), identity strength was measured prior to the main experimental session. The participants worked on the affirmation and belonging subscale of the Multi-Group Ethnic Identity Measure (MEIM, Phinney, 1992). This subscale measures an immigrant's attachment to his or her ethnic group. The results yielded a significant three-way interaction between group (Asian versus Latino), the activation of ethnicity (supposedly eliciting threat among Latinos) and ethnic identity strength. For Latinos, a stereotype threat treatment effect emerged among those with high scores on ethnic identity strength, while the treatment had no influence on the performance of those with low scores.

In a similar vein, one experiment by Schultz et al. (unpublished manuscript, Study 2) revealed a significant interaction between stereotype threat condition and ethnic identity strength. Latino participants with a strong ethnic identity scored lower on the verbal exam in the threat condition. In contrast, higher ethnic identity scores predicted higher verbal exam scores in the control condition. In a follow-up study by Schultz et al. (unpublished manuscript, Study 3) only a main effect for ethnic identity strength was identified, suggesting that this measure was

negatively related to performance across groups. A moderation effect between stereotype threat treatment and identity strength could not be established, the stereotype threat effects were independent of ethnic identity strength.

In the fourth relevant experiment, immigrants from various origins in Austria (most frequent were Kosovo, Bosnia, and Turkey) were exposed to anti-immigrant political ads or neutral political ads (Appel, 2012). The MEIM was used to measure ethnic identity strength after the treatment and the dependent variables. No significant interaction between ethnic identity strength and the treatment was found. However, a simple slope analysis indicated that in this experiment, ethnic identity might have worked as a buffer. For participants low in ethnic identity strength a significant influence of the treatment was observed, whereas participants with high scores on ethnic identity strength were unaffected by the treatment.

In sum, evidence for the role of ethnic identity strength in stereotype threat studies with immigrant samples is sparse and controversial; two experiments observed a significant moderation effect whereas two other studies are inconclusive, one even suggesting that ethnic identity strength can serve as a buffer. It further needs to be noted that, thus far, no study has examined the role of residence culture identity strength for immigrants in situations of stereotype threat.

## Discussion

Historically, many immigrant groups have been faced with negative achievement related stereotypes. In the 1750s, for example, Benjamin Franklin was known for his skepticism regarding German immigrants in Pennsylvania, the “swarthy” “Palatine Boors” (Franklin in Labaree, 1959) who at that time were widely perceived to be lazy, illiterate, and reluctant to assimilate (Feer, 1952). Today, several immigrant groups in most parts of the world underperform in educational settings. Stereotype threat theory posits that negative stereotypes about one’s group can elicit an extra pressure not to fail which leads to cognitive underperformance. Thus, stereotype threat could explain a substantial part of the immigrant achievement gap, one of the arguably most pressing problems for educational research and practice. This meta-analytic review summarized experiments in which immigrants worked on a test of cognitive performance under conditions of low versus high stereotype threat.

Our meta-analysis of 19 independent effect sizes show that the average stereotype threat treatment effect is substantial and significant. This result holds for stereotype threat effects among US samples of Latino background ( $k = 11$ ) as well as among immigrant samples in Europe with various ethnic backgrounds ( $k = 8$ ). All studies conducted in the US were based on young adults (mostly undergraduates). The effects found for adults were larger than for samples of immigrant children and adolescents ( $k = 5$ , all from Europe). Our findings showed that all treatment types (subtle, moderately explicit, blatant) were effectively altering immigrants’ performance. Likewise, studies that used performance measures with a rather strong

learnt knowledge component (GRE-type) and studies that used performance measures with a rather strong fluid intelligence component (memory tasks, RPM-typed tasks) yielded significant stereotype threat effects. Over half of our datasets ( $k = 10$ ) originated from unpublished or yet-to-be-published research. For both types of studies, published or unpublished, substantial and significant average effect sizes were found. A fail-safe analysis (Orwin, 1983; Lipsey and Wilson, 2001) indicated that for the effect size to drop to  $d = 0.2$ , an additional 46 unaccounted for studies were needed, and for the effect size to drop to  $d = 0.1$ , the number of unaccounted for studies amounted to 111.

The comparison between published versus unpublished research and the fail-safe analysis suggest that the sample of studies is not strongly influenced by publication or reporting bias. The funnel plot, however, shows a remarkable pattern of asymmetry: studies with smaller samples tend to yield larger effect sizes in support of stereotype threat than studies with larger samples. There are at least two possible explanations for such an asymmetry. First of all, the corpus of retrievable studies could be subject to a systematic bias in publication and – regarding unpublished research – in accessibility. Publication bias is common in the social sciences, as statistically significant results are more likely to be published than non-significant results, particularly if a small sample study is underpowered. Regarding unpublished research, studies that are underpowered and yield no significant differences are more likely to leave almost no trace that they were ever conducted. Such studies might not be included in a dissertation, they may not be presented at conferences, and so on.

Second, the asymmetrical pattern could be due to heterogeneity among the studies. Studies with larger samples could have a different design or method than studies with smaller samples. For example, small sample studies might be based on individual or small group experimental sessions whereas large sample studies might be based on large-group experimental sessions. The latter procedure, in turn, could increase error variance and therefore decrease the likelihood of identifying significant treatment effects. Due to the fact that few study reports included the number of participants for each experimental session, a systematic analysis of this explanation is impossible. However, our moderator analysis identified the method with which immigrant status was assessed to be an influential factor regarding the studies’ results. The three studies with the largest sample sizes were also among the four studies with the least support for stereotype threat effects. All three studies were among European studies that determined immigrant status with the help of demographic categories, rather than self-identification. For example, Wicherts et al. (2005) used the participants’ parents’ place of birth to decide whether a participant was considered as a member of the residence culture group (Dutch) or the immigrant group. If at least one parent was born outside the Netherlands, the participant would fall into the immigrant category. The authors further report that 17% of those categorized as immigrants self-identified strongly to be Dutch (rather than identifying with Antilles/Suriname, Turkey, or Morocco). Another 16% identified both with the Netherlands

and their immigrant group. As a consequence, the stereotype threat manipulation might have failed to elicit stereotype threat among those who self-identified to be Dutch which amounts to one third of the study sample. In summary, the fact that studies with larger samples yielded less support for stereotype threat effects is likely due to a combination of publication/accessibility bias and heterogeneity of the studies regarding immigrant status assessment.

Our analyses as a whole, including the fail-safe-n analysis and the substantial effect of unpublished research lead to the conclusion that the detrimental effect of stereotype threat on immigrants' performance has been convincingly established by prior research. Therefore, reducing the detrimental impact of stereotype threat for negatively stereotyped immigrants in educational settings appears to be an important objective for research and educators alike.

### Limitations and Future Research Directions

As with all studies, the present review and meta-analysis has limitations that need to be noted. First, our aim in the current meta-analysis was to identify main effects of the stereotype threat treatment. According to stereotype threat theory and research (Schmader et al., 2008), these main effects can be subject to moderation effects identified with the help of additional experimental factors or questionnaires. Due to its key importance in acculturation research we had a closer look on ethnic identity strength as a variable that might affect the magnitude (and direction) of stereotype threat effects. Other moderators were idiosyncratic to single studies. This review and meta-analysis will likely inspire future research rather than dispirit researchers, because exciting questions are still to be answered (cf. Chan and Arvey, 2012). Future research is encouraged to further clarify variations in stereotype threat effects. To name an example, the length of time spent in a country (or generational differences) should be examined more closely. So far, studies within the stereotype threat framework have primarily addressed social groups that have a life-time of exposure to being stereotyped (because of their gender or the color of their skin). The finding that stereotype threat also is found among immigrants raises the question of how much exposure to being stereotyped is necessary for the detrimental effects of stereotypes to become visible, which is a highly important question from a theoretical point of view. Heterogeneity among immigrant groups might be of importance here. While some may be able to merge into the new culture and become less affected by stereotypes over time, for others – due to their merging in a stereotyped subgroup – the susceptibility to stereotype threat may increase (Deaux et al., 2007). The visibility of belonging to an immigrant group might be but one factor that plays a role.

Second, our focus was on performance measures. Although stereotype threat is best known for its influence in test-taking situations, theory and recent research indicate that stereotype threat might affect individuals prior to test-taking (Appel and Kronberger, 2012); that it can lead to short-term (Steele, 1992, 1997; Davies et al., 2002) and long-term disidentification (Osborne and Walker, 2006; Kronberger and Horwath, 2013),

and less successful preparation and learning (Rydell et al., 2010; Appel et al., 2011). Stereotype threat studies on domain identification, career choice or learning are generally rare, and we identified no study on any of these topics that involved immigrants.

Third, the contribution of this meta-analysis regarding the immigrant achievement gap is indirect. The present meta-analysis did not compare the performance of immigrants versus non-immigrants. Still, our results suggest that immigrants' performance varies depending on whether stereotype threat is increased or reduced, which in turn can increase or decrease the distance to non-immigrants' performance.

Fourth, the specific content of stereotypes directed at different groups of immigrants demands for closer inspection, particularly in non-US contexts; the *Stereotype Content Model* (Lee and Fiske, 2006) identifies subgroups and their perception in society along the two dimensions of warmth and competence. This approach seems promising for future attempts to clarify the relationship between stereotype content, status of an immigrant group in society, and the effects of stereotype threat.

Finally, despite some efforts to include as many studies as possible from around the world, our search strategy (focus on English reports, etc.) may have privileged studies on immigration to the US and Europe. There clearly is a need to address the topic in other parts of the world as well.

Investigating stereotype threat among immigrants touches the huge field of research on immigrants' identity and acculturation processes. Regarding questions of identity, there are several points to be made that may inform future research. First, the present research indicates that researchers need to thoroughly reflect about their method to assess immigrant status. Self-identification and (parent's) birthplace methods may at times yield conflicting results. Second, ethnic identity strength has been conceived as a one- or two-dimensional concept, including the sub-factors commitment and exploration (e.g., Phinney, 1992). In future research it could be useful to consider additional dimensions and facets, for example, private regard of one's ethnic culture or remigration prospects. So far, this aspect has not been sufficiently addressed in stereotype threat research. Third, there is a research history of describing immigrant identity along the lines of both ethnic identity and residence culture identity (sometimes addressed as 'national identity'). Future research on stereotype threat among immigrants is advised to examine both identity dimensions. Possibly, individuals with a strong residence culture identity are less affected by potentially threatening cues, independent of their ethnic identity strength. Immigrants with a strong residence culture identity might at times switch between identities, as they can be regarded *bicultural*, being able to draw from resources of two (cultural) backgrounds. There is evidence that it is possible to activate a particular identity in people with a bicultural identity (cf. Hong et al., 2000). Consequently, people then think, feel, and act according to the activated identity. If this is the case, active cultural frame switching could protect bicultural immigrants in situations of stereotype threat (Shih et al., 1999). As a consequence, stereotype threat vulnerability might be decreased due to a strong identification with the residence culture.

## Implications

Our review and meta-analysis suggests that stereotype threat is a psychological predicament with detrimental consequences on immigrants' performance. Thus, reducing the negative influence of stereotype threat could help closing the achievement gap. Stronger cognitive performance and educational success can increase the opportunities of immigrants to participate in their resident societies. Moreover, closing the immigrant achievement gap could be one important key to the future prosperity of societies. According to the (World Economic Forum [WEF], 2011) many countries in the Northern hemisphere will soon be faced with shortages of proficient workforce due to an aging population and insufficient education. According to this economic data, Western Europe, for example, is required to add over 45 million people to its talent base by 2030 in order to sustain economic growth. As well-educated personnel is becoming increasingly scarce, companies and countries "will need to be creative in identifying under-utilized or underdeveloped pools of talent" (p. 19).

Previous research on gender and race identified strategies to reduce the negative influence of stereotype threat (e.g., for an overview see Gehlbach, 2010; Yeager and Walton, 2011; Cohen et al., 2012). A main approach is reducing the salience of the negative link between belonging to a minority group and learning and performance at school. As our meta-analytic findings demonstrate, even subtle cues, like asking negatively stereotyped immigrant students about their ethnic background prior to a task, can decrease performance to the same extent as blatant cues, such as stating the cognitive inferiority of a group. This aspect is crucial for practical implications in school or work environments, where immigrants might constantly experience the fear to underperform due to subtle stereotype activating cues, even if blatant cues (e.g., use of discriminatory or sexist language) are largely frowned upon and regarded as politically incorrect. The resulting actual underperformance might then reinforce the negative stereotype (self-fulfilling prophecy), which makes it hard to break the cycle. Moreover, valuing diversity and increasing the visibility of minority group members who may be fellow students or teachers decreases stereotype threat effects (e.g., Purdie-Vaughns et al., 2008;

Carrell et al., 2010). Stereotyped students can further profit from successful role models (e.g., McIntyre et al., 2003), but perceived dissimilarities between oneself and the role model might reduce the influence or even yield detrimental effects (e.g., Cheryan et al., 2011; Asgari et al., 2012). An intriguing line of research showed that brief classroom interventions that highlight students' core personal values can improve the performance of negatively stereotyped students (e.g., Cohen et al., 2006, 2009; Miyake et al., 2010; cf., Yeager and Walton, 2011).

The attractiveness of the social-psychological strategies to increase school performance lies in the fact that small and easy to implement interventions can result in considerable improvements for those concerned. To date, very little research on these strategies has focused on immigrant students. However, our review on stereotype threat suggests that these strategies, applied to the group of immigrants, could reduce the detrimental effects of stereotype threat and could contribute to a better achievement of immigrant students. In line with this assumption, a field experiment on values affirmation provides initial evidence on the effectiveness of these strategies for immigrant students (Sherman et al., 2013). In addition to academic success, research on stereotype threat spillover (Inzlicht and Kang, 2010; Aronson et al., 2013) suggests that interventions to combat stereotype threat might effectively reduce immigrant-majority member gaps in fields such as health or aggression. In summary our analyses suggest that immigrant students in many parts of the world are confronted with a stereotype threat that endangers them to perform below their potential. However, our analyses also suggest that there is considerable need for further research to better understand how this threat actually works and how it can be mitigated.

## Acknowledgments

We are indebted to all authors who provided us with information from unpublished studies. Preparation of this manuscript was supported by the ÖNB Anniversary Fund (Austrian Central Bank, project number: 14994).

## References

- Appel, M. (2012). Anti-immigrant propaganda by radical right parties and the intellectual performance of adolescents. *Polit. Psychol.* 33, 483–493. doi: 10.1111/j.1467-9221.2012.00902.x
- Appel, M., and Kronberger, N. (2012). Stereotype threat and the achievement gap: stereotype threat prior to test taking. *Educ. Psychol. Rev.* 24, 609–635. doi: 10.1007/s10648-012-9200-4
- Appel, M., Kronberger, N., and Aronson, J. (2011). Stereotype threat impedes ability building: effects on the test preparation of women in science and technology. *Eur. J. Soc. Psychol.* 41, 904–913. doi: 10.1002/ejsp.835
- Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: the moderating role of ethnic identification. *Cult. Divers. Ethnic. Minor. Psychol.* 16, 94–98. doi: 10.1037/a0017564
- Armenta, B. E., and Hunt, J. S. (2009). Responding to societal devaluation: effects of perceived personal and group discrimination on the ethnic group identification and personal self-esteem of Latino/Latina adolescents. *Group Process. Intergr. Relat.* 12, 23–39. doi: 10.1177/1368430208098775
- Aronson, J., Burgess, D., Phelan, S. M., and Juarez, L. (2013). Unhealthy interactions: the role of stereotype threat in health disparities. *Am. J. Public Health* 103, 50–56. doi: 10.2105/AJPH.2012.300828
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., and Brown, J. (1999). When white men can't do math: necessary and sufficient factors in stereotype threat. *J. Exp. Soc. Psychol.* 35, 29–46. doi: 10.1006/jesp.1998.1371
- Asgari, S., Dasgupta, N., and Stout, J. G. (2012). When do counterstereotypic ingroup members inspire versus deflate? The effect of successful professional women on young women's leadership self-concept. *Pers. Soc. Psychol. Bull.* 38, 370–383. doi: 10.1177/0146167211431968
- Beilock, S. L., Rydell, R. J., and McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spill over. *J. Exp. Psychol. Gen.* 136, 256–276. doi: 10.1037/0096-3445.136.2.256

- Benet-Martinez, V., and Haritatos, J. (2005). Bicultural identity integration (BII): components and psychosocial antecedents. *J. Pers.* 73, 1015–1050. doi: 10.1111/j.1467-6494.2005.00337.x
- Berjot, S., Roland-Levy, C., and Girault-Lidvan, N. (2011). Cognitive appraisals of stereotype threat. *Psychol. Rep.* 108, 585–598. doi: 10.2466/04.07.21.PR0.108.2.585-598
- Berry, J. W. (1990). "Psychology of acculturation: understanding individuals moving between cultures," in *Applied Cross-Cultural Psychology*, ed. R. W. Brislin (Thousand Oaks, CA: Sage), 232–253.
- Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Appl. Psychol. Int. Rev.* 46, 5–34. doi: 10.1080/026999497378467
- Brown, R. P., and Pinel, E. C. (2003). Stigma on my mind: individual differences in the experience of stereotype threat. *J. Exp. Soc. Psychol.* 39, 626–633. doi: 10.1016/S0022-1031(03)00039-8
- Carrell, S. E., Page, M. P., and West, J. E. (2010). Sex and science: how professor gender perpetuates the gender gap. *Q. J. Econom.* 125, 1101–1144. doi: 10.3386/w14959
- Cattell, R. B. (1940). A culture-free intelligence test. *J. Educ. Psychol.* 31, 161–179. doi: 10.1037/h0059043
- Chan, M. E., and Arvey, R. D. (2012). Meta-analysis and the development of knowledge. *Perspect. Psychol. Sci.* 7, 79–92. doi: 10.1177/1745691611429355
- Chatard, A., Selimbegović, L., Konan, P., and Mugny, G. (2008). Performance boosts in the classroom: stereotype endorsement and prejudice moderate stereotype lift. *J. Exp. Soc. Psychol.* 44, 1421–1424. doi: 10.1016/j.jesp.2008.05.004
- Chateignier, C., Dutrévis, M., Nugier, A., and Chekroun, P. (2009). French-Arab students and verbal intellectual performance: do they really suffer from a negative intellectual stereotype? *Eur. J. Psychol. Educ.* 24, 219–234. doi: 10.1007/BF03173013
- Cheryan, S., and Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: the psychological hazards of "model minority" status. *Psychol. Sci.* 11, 399–402. doi: 10.1111/1467-9280.00277
- Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., and Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM? *Soc. Psychol. Pers. Sci.* 2, 656–664. doi: 10.1177/1948550611405218
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, G. L., Garcia, J., Apfel, N., and Master, A. (2006). Reducing the racial achievement gap: a social-psychological intervention. *Science* 313, 1307–1310. doi: 10.1126/science.1128317
- Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., and Brzustoski, P. (2009). Recursive processes in self-affirmation: intervening to close the minority achievement gap. *Science* 324, 400–403. doi: 10.1126/science.1170769
- Cohen, G. L., Purdie-Vaughns, V., and Garcia, J. (2012). "An identity threat perspective on intervention," in *Stereotype Threat: Theory, Process, and Application*, eds M. Inzlicht and T. Schmader (New York, NY: Oxford University Press), 280–296.
- Davies, P. G., Spencer, S. J., Quinn, D. M., and Gerhardstein, R. (2002). Consuming images: how television commercials that elicit stereotype threat can restrain women academically and professionally. *Pers. Soc. Psychol. Bull.* 28, 1615–1628. doi: 10.1177/014616702237644
- Deaux, K. (1996). "Social identification," in *Social Psychology: Handbook of Basic Principles*, eds E. T. Higgins and A. W. Kruglanski (New York, NY: Guilford Press), 777–798.
- Deaux, K., Bikmen, N., Gilkes, A., Ventuneac, A., Joseph, Y., Payne, R., et al. (2007). Becoming American: stereotype threat effects in Afro-Caribbean immigrant groups. *Soc. Psychol. Q.* 70, 384–404. doi: 10.1177/01902725070700408
- Enesco, I., Navarro, A., Paradelo, I., and Guerrero, S. (2005). Stereotypes and beliefs about different ethnic groups in Spain. A study with Spanish and Latin American children living in Madrid. *J. Appl. Dev. Psychol.* 26, 638–659. doi: 10.1016/j.appdev.2005.08.009
- Eriksson, K., and Lindholm, T. (2007). Making gender matter: the role of gender-based expectancies and gender identification on women's and men's math performance in Sweden. *Scand. J. Psychol.* 48, 329–338. doi: 10.1111/j.1467-9450.2007.00588.x
- Feer, R. A. (1952). Official use of the German language in Pennsylvania. *Magazine His. Biogr.* 76, 394–405.
- Flanagan, J. (2009). *The Effects of Stereotype Threat on African-American and Hispanic Participants on Academic and Non-Academic Tasks in Workplace Settings*. dissertation, Texas A&M University, College Station, TX.
- Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., and Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Dev. Psychol.* 49, 1886–1897. doi: 10.1037/a0031412
- Gehlbach, H. (2010). The social side of school: why teachers need social psychology. *Educ. Psychol. Rev.* 22, 349–362. doi: 10.1007/s10648-010-9138-3
- Good, C., Aronson, J., and Inzlicht, M. (2003). Improving adolescents' standardized test performance: an intervention to reduce the effects of stereotype threat. *J. Appl. Dev. Psychol.* 24, 645–662. doi: 10.1016/j.appdev.2003.09.002
- Gonzales, P. M., Blanton, H., and Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of latino women. *Pers. Soc. Psychol. Bull.* 28, 659–670. doi: 10.1177/0146167202288010
- Hatton, T. J., and Williamson, J. G. (1994). What drove the mass migrations from Europe in the late nineteenth century? *Popul. Dev. Rev.* 20, 1–27. doi: 10.2307/2137600
- Headrick, T. C. (2010). *Statistical Simulation: Power Method Polynomials and other Transformations*. Boca Raton, FL: Chapman & Hall/CRC.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *J. Educ. Behav. Statist.* 6, 107–128. doi: 10.3102/10769986006002107
- Ho, A. K., and Sidanius, J. (2010). Preserving positive identities: public and private regard for one's ingroup and susceptibility to stereotype threat. *Group Process. Intergr. Relat.* 13, 55–67. doi: 10.1177/1368430209340910
- Hollis-Sawyer, L. A., and Sawyer, T. P. Jr. (2008). Potential stereotype threat and face validity effects on cognitive-based test performance in the classroom. *Educ. Psychol.* 28, 291–304. doi: 10.1080/01443410701532313
- Hong, Y.-Y., Morris, M., Chiu, C.-Y., and Benet-Martinez, V. (2000). Multicultural minds: a dynamic constructivist approach to culture and cognition. *Am. Psychol.* 55, 709–720. doi: 10.1037/0003-066X.55.7.709
- Howard, K. E., and Anderson, K. A. (2010). Stereotype threat in middle school: the effects of prior performance on expectancy and test performance. *Middle Grades Res. J.* 5, 119–137.
- Institute National de la Statistique et des Etudes Economiques [INSEE]. (2013). *Répartition des Immigrés par Pays de Naissance*. Available at: [http://www.insee.fr/fr/themes/tableau.asp?reg\\_id=0&ref\\_id=immigrespaysnaiss](http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=immigrespaysnaiss)
- Inzlicht, M., and Kang, S. K. (2010). Stereotype threat spillover: how coping with threats to social identity affects, aggression, eating, decision-making, and attention. *J. Pers. Soc. Psychol.* 99, 467–481. doi: 10.1037/a0018951
- Inzlicht, M., and Schmader, T. (eds). (2012). *Stereotype Threat: Theory, Process, and Application*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199732449.001.0001
- Jäckle, N. (2008). *Die ethnische Hierarchie in Deutschland und die Legitimierung der Ablehnung und Diskriminierung Ethnischer Minoritäten* [The Ethnic Hierarchy in Germany and the Legitimization of Rejecting and Discriminating Ethnic Minorities]. Doctoral dissertation, University of Marburg, Marburg.
- Kahraman, B., and Knoblich, G. (2000). Stechen statt sprechen: valenz und aktivierbarkeit von stereotypen über türken. *Z. Sozialpsychol.* 31, 31–43. doi: 10.1024//0044-3514.31.1.31
- Klein, O., Pohl, S., and Ndagijimana, C. (2007). The influence of intergroup comparisons on Africans' intelligence test performance in a job selection context. *J. Psychol.* 141, 453–467. doi: 10.3200/JRLP.141.5.453-468
- Kronberger, N., and Horwath, I. (2013). The ironic costs of performing well: grades differentially predict male and female dropout from engineering. *Basic Appl. Soc. Psychol.* 35, 534–546. doi: 10.1080/01973533.2013.840629
- Labaree, L. W. (ed.). (1959). *The Papers of Benjamin Franklin*. New Haven, MA: Yale University Press.

- Lee, T. L., and Fiske, S. T. (2006). Not an outgroup, not yet an ingroup: immigrants in the stereotype content model. *Int. J. Intercult. Relat.* 30, 751–768. doi: 10.1016/j.ijintrel.2006.06.005
- Liebkind, K. (2001). “Acculturation,” in *Blackwell Handbook of Social Psychology: Intergroup Processes*, eds R. Brown and S. Gaertner (Oxford: Blackwell), 386–406.
- Lipsey, M. W., and Wilson, D. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
- Logel, C., Peach, J. M., and Spencer, S. J. (2012). “Threatening gender and race: different manifestations of stereotype threat,” in *Stereotype Threat: Theory, Process, and Application*, eds M. Inzlicht and T. Schmader (New York, NY: Oxford University Press), 159–172.
- McFarland, L. A., Lev-Arey, D. M., and Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Hum. Perform.* 16, 181–205. doi: 10.1207/S15327043HUP1603\_2
- McIntyre, R. B., Paulson, R. M., and Lord, C. G. (2003). Alleviating women’s mathematics stereotype threat through salience of group achievements. *J. Exp. Soc. Psychol.* 39, 83–90. doi: 10.1016/S0022-1031(02)00513-9
- Miyake, A., Smith-Kost, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., and Ito, T. A. (2010). Reducing the gender achievement gap in college science: a classroom study of values affirmation. *Science* 330, 1234–1237. doi: 10.1126/science.1195996
- Nadler, J. T., and Clark, M. H. (2011). Stereotype threat: a meta-analysis comparing African Americans to Hispanic Americans. *J. Appl. Soc. Psychol.* 41, 872–890. doi: 10.1111/j.1559-1816.2011.00739.x
- Nguyen, A. D., and Benet-Martínez, V. (2010). “Multicultural identity: what it is and why it matters,” in *The Psychology of Social and Cultural Diversity*, ed. R. J. Crisp (Hoboken, NJ: Wiley-Blackwell), 87–114. doi: 10.1002/9781444325447.ch5
- Nguyen, H.-H. D., and Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *J. Appl. Psychol.* 93, 1314–1334. doi: 10.1037/a0012702
- OECD. (2010). *Closing the Gap for Immigrant Students: Policies, Practice and Performance*. Paris: OECD Publishing. doi: 10.1787/9789264075788-en
- OECD. (2013). *International Migration Outlook 2013*. Paris: OECD Publishing. doi: 10.1787/migr\_outlook-2013-en
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *J. Educ. Statist.* 8, 157–159. doi: 10.2307/1164923
- Osborne, J., and Walker, C. (2006). Stereotype threat, identification with academics, and withdrawal from school: why the most successful students of color might be most likely to withdraw. *Educ. Psychol.* 26, 563–577. doi: 10.1080/01443410500342518
- Pellegrini, A. (2005). *The Impact of Stereotype Threat on Intelligence Testing in Hispanic Females*. dissertation. Carlos Albizu University, Miami, FL.
- Phinney, J. S. (1992). The multigroup ethnic identity measure: a new scale for use with diverse groups. *J. Adolesc. Res.* 7, 156–176. doi: 10.1177/074355489272003
- Phinney, J. S., Horenczyk, G., Liebkind, K., and Vedder, P. (2001). Ethnic identity, immigration, and well-being: an interactional perspective. *J. Soc. Issues* 57, 493–510. doi: 10.1111/0022-4537.00225
- Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Dittmann, R., and Crosby, J. R. (2008). Social identity contingencies: how diversity cues signal threat or safety for African Americans in mainstream institutions. *J. Pers. Soc. Psychol.* 94, 615–630. doi: 10.1037/0022-3514.94.4.615
- Raven, J. C. (1958). *Standard Progressive Matrices*. London: Lewis & Co. Ltd.
- Romero, A. J., and Roberts, R. E. (2003). The impact of multiple dimensions of ethnic identity on discrimination and adolescents’ self-esteem. *J. Appl. Soc. Psychol.* 33, 2288–2305. doi: 10.1111/j.1559-1816.2003.tb01885.x
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychol. Bull.* 86, 638–641. doi: 10.1037/0033-2909.86.3.638
- Rydell, R. J., Rydell, M. T., and Boucher, K. L. (2010). The effect of negative performance stereotypes on learning. *J. Pers. Soc. Psychol.* 99, 883–896. doi: 10.1037/a0021139
- Salinas, M. F. (1998). *Stereotype Threat: The Role of Effort Withdrawal and Apprehension on the Intellectual UnderPerformance of Mexican Americans*. Dissertation, University of Texas, Austin, TX.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women’s math performance. *J. Exp. Soc. Psychol.* 38, 194–201. doi: 10.1006/jesp.2001.1500
- Schmader, T., and Beilock, S. (2012). “An integration of processes that underlie stereotype threat,” in *Stereotype Threat: Theory, Process, and Application*, eds M. Inzlicht and T. Schmader (New York, NY: Oxford University Press), 34–50.
- Schmader, T., and Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *J. Pers. Soc. Psychol.* 85, 440–452. doi: 10.1037/0022-3514.85.3.440
- Schmader, T., Johns, M., and Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychol. Rev.* 115, 336–356. doi: 10.1037/0033-295X.115.2.336
- Shapiro, J. S., and Neuberg, S. L. (2007). From stereotype threat to stereotype threats: implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Pers. Soc. Psychol. Rev.* 11, 107–130. doi: 10.1177/1088868306294790
- Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., et al. (2013). Deflecting the trajectory and changing the narrative: how self-affirmation affects academic performance and motivation under identity threat. *J. Pers. Soc. Psychol.* 104, 591–618. doi: 10.1037/a0031495
- Shih, M., Ambady, N., Richeson, J. A., Fujita, K., and Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *J. Pers. Soc. Psychol.* 83, 638–647. doi: 10.1037/0022-3514.83.3.638
- Shih, M., Pittinsky, T. L., and Ambady, N. (1999). Stereotype susceptibility: identity salience and shifts in quantitative performance. *Psychol. Sci.* 10, 80–83. doi: 10.1111/1467-9280.00111
- Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women’s math performance. *J. Exp. Soc. Psychol.* 35, 4–28. doi: 10.1006/jesp.1998.1373
- Steele, C. M. (1992). Race and schooling of black Americans. *Atlantic Monthly* 269, 68–78.
- Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *Am. Psychol.* 52, 613–629. doi: 10.1037/0003-066X.52.6.613
- Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *J. Pers. Soc. Psychol.* 69, 797–811. doi: 10.1037/0022-3514.69.5.797
- Steele, C. M., Spencer, S. J., and Aronson, J. (2002). “Contending with images of one’s group: the psychology of stereotype and social identity threat,” in *Advances in Experimental Social Psychology*, ed. M. Zanna (San Diego, CA: Academic Press).
- Stone, J. (2002). Battling doubt by avoiding practice: the effects of stereotype threat on self-handicapping in white athletes. *Pers. Soc. Psychol. Bull.* 28, 1667–1678. doi: 10.1177/014616702237648
- Stündendorfer, A. (2006). *Stereotype Threat – Eine Bedrohung für Türkische Schüler an Deutschen Grundschulen?* [Stereotype threat – a threat for Turkish students at German elementary schools?] Diploma thesis, Friedrich-Alexander University Erlangen-Nürnberg, Nürnberg.
- Taylor, V. J., and Walton, G. M. (2011). Stereotype threat undermines academic learning. *Pers. Soc. Psychol. Bull.* 37, 1055–1067. doi: 10.1177/0146167211406506
- Tenenbaum, H. R., and Ruck, M. D. (2007). Are teachers’ expectations different for racial minority than for European American students? A meta-analysis. *J. Educ. Psychol.* 99, 253–273. doi: 10.1037/0022-0663.99.2.253
- Umaña-Taylor, A. J., and Updegraff, K. A. (2007). Latino adolescents’ mental health: exploring the interrelations among discrimination, ethnic identity, cultural orientation, self-esteem, and depressive symptoms. *J. Adolesc.* 30, 549–567. doi: 10.1016/j.adolescence.2006.08.002
- Umaña-Taylor, A. J., Wong, J. J., Gonzales, N. A., and Dumka, L. E. (2012). Ethnic identity and gender as moderators of the association between discrimination and academic adjustment among Mexican-origin adolescents. *J. Adolesc.* 35, 773–786. doi: 10.1016/j.adolescence.2011.11.003
- US Department of Homeland Security. (2013). *Immigration Statistics*. Available at: <http://www.dhs.gov/immigration-statistics>
- Verkuyten, M. J. A. M., and Kinket, B. M. (1999). The relative importance of ethnicity: ethnic categorization among older children. *Int. J. Psychol.* 34, 107–118. doi: 10.1080/002075999400005

- Walton, G. M., and Cohen, G. L. (2003). Stereotype lift. *J. Exp. Soc. Psychol.* 39, 456–467. doi: 10.1016/S0022-1031(03)00019-2
- Walton, G. M., and Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychol. Sci.* 20, 1132–1139. doi: 10.1111/j.1467-9280.2009.02417.x
- Waters, M. C. (1999). *Black Identities: West Indian Immigrant Dreams and American Realities*. New York: Russell Sage Foundation.
- Wicherts, J. M., Dolan, C. V., and Hessen, D. J. (2005). Stereotype threat and group differences in test performance: a question of measurement invariance. *J. Pers. Soc. Psychol.* 89, 696–716. doi: 10.1037/0022-3514.89.5.696
- Wilson, D. B. (2012). *Practical Meta-analysis Effect Size Calculator*. Available at: <http://gemini.gmu.edu/cebcp/EffectSizeCalculator/index.html>
- World Economic Forum [WEF]. (2011). *World Economic Forum in Collaboration with The Boston Consulting Group . Global Talent. Risk – Seven Responses*. Geneva: WEF.
- Yeager, D. S., and Walton, G. M. (2011). Social-psychological interventions in education: they're not magic. *Rev. Educ. Res.* 81, 267–301. doi: 10.3102/0034654311405999

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Appel, Weber and Kronberger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.