Quality & Quantity
https://doi.org/10.1007/s11135-022-01512-8

The conundrum of gender-science stereotypes: a review and discussion of measurements

Elena De Gioannis[1]

Accepted: 4 August 2022

Abstract
Stereotypes do not have a unique definition, being mostly considered a generalized belief on the quality and characteristics of members of specific groups or social categories. Hence, various scales and measurements have been proposed to assess the endorsement of beliefs on the association of gender and scientific/language-related skills. The aim of the paper was to summarize, compare and discuss those measures, distinguishing between explicit, implicit and indirect measures. The review of the literature highlighted a huge but unrecognized heterogeneity in the constructs of gender stereotypes, especially for explicit measures. This can hamper findings comparability, reduce scales' validity, affect the correlation between implicit and explicit measurements, and bias their interpretations due to ambiguous terminologies.

Keywords Gender stereotypes · Gender-science stereotypes · Literature review · Gender

1 Introduction

In the last years, gender stereotypes have been acknowledged a crucial role in determining and contributing to the underrepresentation of women in STEM (Science, Technology, Engineering, and Math). Several studies have concurrently attempted to test their effect on girls and women's aspirations, performance, interests and sense of belongingness in this field (Cundiff et al. 2013; Kiefer and Sekaquaptewa 2007a, b; Lane et al. 2012; Nosek and Smyth 2011; Reuben et al. 2014). On the other hand, others have tested how these gender stereotypes can be effectively reduced. Interventions with this purpose include, for instance, exposing girls to counterstereotypical role models (Betz and Sekaquaptewa 2012; Dasgupta and Asgari 2004; McIntyre et al. 2003) and making them aware of the detrimental influence of stereotypes' endorsement (Farrell et al. 2020; Jackson et al. 2014; Johns et al. 2005).

Despite being similar in the addressed issue, these studies show great variability in the target population (children, adolescents, adults), the variables of interest, the setting

✉ Elena De Gioannis
elena.degioannis@unimi.it

[1] Department of Social and Political Sciences, University of Milan, Via Conservatorio 7, 20122 Milan, Italy

Published online: 26 August 2022

(laboratory or field) and the research design. This heterogeneity may explain the lack of a unique and shared scale to measure gender-science stereotypes. Conversely, scales of this type exist for stereotypes on other gender-related issues, e.g., the Attitudes toward Women Scale (Spence and Helmreich 1972) for gender roles and the Ambivalent Sexism Inventory (Glick and Fiske 1996) for sexism.

Furthermore, stereotypes do not have a unique definition and the lack of a univocal meaning may also explain this heterogeneity in measurements. In their review of instruments for gender roles, McHugh and Frieze (1997) blamed the proliferation of scales that make any comparison difficult. While the existence of several and different instruments is a problem, the multi-facet nature of gender-related characteristics and gender-belief systems would require avoiding single measurements. Indeed, gender stereotypes address several issues, i.e., traits, attitudes, interests, cognitive skills, family roles and occupations (Hentschel et al. 2019; Six and Eckes 1991).

Given the existence of such a multitude of indicators, previous articles reviewed and synthesized scales and other types of measures used to assess gender stereotypes on roles (Beere 1990; McHugh and Frieze 1997), traits and attitudes (Kite et al. 2008; Smiler 2004) but, to the best of our knowledge, only Zitelny et al. (2017) grouped studies on gender-science stereotypes and listed the instruments used. However, their aim was not to review these instruments, but rather to focus the attention on the correlation between implicit and explicit measures. There is a need for filling this gap and comparing measures of stereotypes on gender and STEM. This is the first aim of this paper.

The second part of the article will discuss potential consequences deriving from the heterogeneity of these instruments and some of their limitations. Expanding on the problem of findings comparison argued by McHugh and Frieze (1997), it is suggested here that the proliferation of scales affects also the correlation between implicit and explicit measures (Zitelny et al. 2017). As regards potential limitations of existing instruments, it is argued that while the focus of current studies is typically on math, also science and reading should be investigated as they are affected by stereotypes as well. Moreover, since some questions behind measurements are often too generic so leaving space for interpretability, there is a need for understanding how this can compromise the interpretation of final scores.

Once acknowledged the increasing interest in gender stereotypes and the need to understand the manyfold role they play in STEM, this study could be beneficial for research on the theme in two ways. On the one hand, researchers have access to a general overview of the instruments available to test gender stereotypes. The summary of results facilitates the identification of instruments and their psychometric soundness while showing also their diffusion in previous studies. This could help reduce the tendency to create new ad hoc measures. On the other hand, the discussion about the instruments' limitations set the ground for a refinement in the measurement of gender stereotypes and suggests an unexplored field of research on the theme.

2 Literature review

2.1 Stereotypes

The word 'stereotype' was first used by the journalist Lippman (1922) to indicate general cognitive structures that serve as mental pictures of social groups. However, since then, the

meaning and definition of stereotypes have changed and evolved (for a review, see Schneider 2005).

Several definitions of stereotypes exist, differing in whether they describe stereotypes as inaccurate, consider stereotypes disagreeable in both the formation process and the consequences and represent stereotypes as shared among people or as individual beliefs (Schneider 2005). To mention just a few, stereotypes were defined as 'Beliefs and opinions about the characteristics, attributes and behaviours of members of various groups' (Hilton and von Hippel 1996, p. 240), as 'Both positive and negative beliefs or overgeneralizations about the attributes of a group and its members' (Marx and Ko 2012, p. 160), and more recently as 'General expectations about members of particular social groups […] that leads people to overemphasize differences between groups and underestimate variations within groups' (Ellemers 2018).

The social groups affected by stereotypes are various. Early research focused on stereotypes of race and ethnicity, while, starting from the 1970s and 1980s, the widespread interest in the discrimination against women led to an expansion in research on gender stereotypes (Schneider 2005). There are several beliefs on gender differences, ranging from characteristics to roles, and, consequently, gender stereotypes consist of multiple components. This paper focuses on the stereotypical belief that women and men would differ in their mathematical and scientific abilities, with men traditionally considered to outperform women in STEM.

2.2 Gender-science stereotypes

The belief that women would perform poorly in STEM and, conversely, that STEM would be the natural domain of men traditionally derived from the unfounded conviction that women and men's brains differ, the latter being more apt to logical thinking (Kersey et al. 2019). Furthermore, the observation of women's underrepresentation in STEM (Eagly and Wood, 2012) and the association of success in these fields with being agentic — a characteristic traditionally attributed to men (Sczesny et al. 2018) — contributed to the reinforcement of beliefs on gender differences not only in abilities but also in interests and aptitudes (e.g., Plante et al. 2009).

It is still unclear when boys and girls start endorsing gender-science stereotypes. On the one hand, to a certain age children tend to consider their gender the smartest (Grow et al. 2016) — ingroup favouritism — on the other hand, in some studies there was evidence that children associated math with boys by the age of six (Master et al. 2017; Tomasetto et al. 2012).

In the last years, research on gender stereotypes in this context has deeply increased and numerous studies found evidence of the detrimental effect of stereotypical beliefs on women in STEM. To mention some, Cundiff et al. (2013) found that among college students, women endorsing stronger gender-science stereotypes had weaker science identification and, in turn, weaker science career aspirations. Kiefer and Sekaquaptewa (2007a) found a negative association between stereotypes and performance. Female students with low implicit gender stereotypes performed better in a calculus course in college compared to those with stronger implicit stereotypes. Finally, Nosek and Smyth (2011) found that stronger implicit gender-science stereotypes predicted women's higher negativity toward math, lower participation in STEM, and worse achievement in math.

The relevance of addressing the issue of stereotypes is widely recognized, to the extent that the Committee on the Elimination of Discrimination against Women (CEDAW) — an

international human rights treaty — regulate states' obligations to address stereotypes and stereotyping affecting women (Cusack 2013). Given these premises, it is not difficult to recognize the importance of using valid instruments, especially, when testing strategies that can potentially be applied on a larger scale.

2.3 Instruments to measure gender-science stereotypes

In the context of stereotypes on gender and STEM, it is not possible to identify a widely adopted instrument assessing stereotypes' endorsement. A subscale of the Fennema-Sherman Mathematics Attitudes Scales (Fennema and Sherman 1976), the Mathematics as a male domain scale, could have fulfilled this role. However, despite further refinement and validations of this scale (Leder and Forgasz 2002) researchers tended to create new instruments, with fewer items and shaped on the aim of their study, rather than adopting the existing scales. This resulted in a proliferation of heterogeneous measurements.

The lack of such scales has several negative consequences. As observed by McHugh and Frieze (1997, p. 3) for gender roles, 'When each researcher […] develops his/her own scale, it becomes increasingly difficult to make comparisons across studies, across samples, across cultures and over time. It is unlikely that each researcher has developed a valid and reliable measure, and even more unlikely that each is measuring a unique, enduring, and important construct'. The existence of multiple instruments measuring the same constructs requires and justifies reviews summarizing and reporting them. While several reviews were published on gender-role attitudes and sexism (Beere 1990; Kite et al. 2008; McHugh and Frieze 1997; Smiler 2004), to our knowledge there is a lack of reviews on gender-science stereotypes.

A partial summary of these instruments is reported in the appendix of the study by Zitelny et al. (2017). They aimed to analyse the correlation between implicit, i.e., the Implicit Association Test (IAT), and explicit measures of gender-science stereotypes to suggest that the former should not be interpreted as a counterpart of the latter. In the appendix, they summarized twenty-four studies in which both explicit and implicit measures were used and reported the correlation between the two. The authors suggested that the observed heterogeneity in the correlation between the two measures may be due to the use of different self-reported instruments.

They distinguished between self-reported beliefs (about natural ability, natural interest and prevalence), and self-reported association, i.e., the extent to which participants associated science with males versus females, and liberal arts with females versus males. They then discussed the relationship between the two and the IAT scores. Results indicated that, among beliefs and self-reported association, the latter is the one that correlates with the IAT the most. This suggests that 'the IAT taps into different constructs than those tapped by the explicit measures used in research on the gender-science stereotype' (Zitelny et al. 2017, p. 6). Consequently, the authors suggested that a distinction among constructs of stereotypes and a more specific choice of one over the other may be relevant when both explicit and implicit measurements are used.

Expanding the summary table in Zitelny et al. (2017), the current study reports a comprehensive overview of papers including an instrument of gender-science stereotypes' endorsement. More specifically, we included explicit, implicit and indirect instruments (Whitley and Kite 2016). The first group refers to instruments based on participants' self-reports, the second group to those measuring the mental association among concepts and the third one to instruments which, as explicit measurements, asked participants about

opinions or beliefs. However, unlike previous ones, in the latter, concepts are only indirectly linked to gender stereotypes.

2.4 Complementary gender stereotypes

The distinction between implicit and explicit measures is not limited to whether they assess automatic or self-reported beliefs. Implicit indicators, e.g., the IAT, mostly provide a final score computed as the difference between the gender-STEM and the gender-humanities automatic associations. As such, 'the IAT is limited to measuring the relative strengths of pairs of associations rather than absolute strengths of single associations. In practice, however, the IAT can nevertheless be effectively used because many socially significant categories form complementary pairs' (Greenwald and Farnham 2000, p. 1023). Conversely, explicit instruments mostly ask participants their opinion about the single gender-STEM association.

Similarly to gender stereotypes on math-related skills, gender stereotypes on reading skills traditionally attribute higher reading skills to females compared to males. Several studies have assessed the diffusion of this stereotype across countries. To mention few, Dwyer (1974) found that girls and boys in the U.S. tended to describe reading as a feminine activity throughout the school years until grade 12. More recently, Retelsdorf et al. (2015) assessed German teachers' endorsement of stereotypes on girls performing better in reading tasks than boys.

Being this distinction between 'complementary stereotypes' (Jost and Kay 2005) potentially relevant when comparing and discussing instruments of stereotypes (Gilbert et al. 2015), here we also reported when and how studies investigating gender-science stereotypes included a measure of the gender-humanities association. Note that here both 'humanities' and 'language/reading' terms are used when referring to this complementary association. This is because studies investigating the association between gender and careers/majors usually refer to humanities (or liberal arts), while those investigating the association between gender and abilities refer to language-related, writing or reading skills (see Tables 1 and 2). Finally, here, only STEM disciplines were considered, as studies on this theme often focus on these four areas. However, note that gender inequality persists also in other science-related disciplines where women's representation has progressively increased (Begeny et al. 2020).

2.5 Methodology

Figure 1 summarizes the procedure followed for the selection of studies. First, a combination of key terms (gender AND stereotyp* AND (STEM OR math* OR scien* OR engineer* OR techn*) AND (instrument* OR measur*)) was searched in relevant databases (Web of Science, PsycINFO, Scopus) in January 2021 and arranged according to the rules of the databases. After a screening of the reports resulting from this first stage, relevant references cited in those reports were then screened. This second stage of the selection process was pivotal to backtrack to the source who proposed the instrument in the first place.

To be eligible for inclusion, studies should have used an instrument to measure the endorsement of stereotypes on gender and STEM-related domains. The eligibility criteria were left wide on purpose, as one of the aims of this review is to highlight the heterogeneity in instruments and how gender stereotypes are described in the studies. Consequently, all studies stating to measure gender stereotypes in STEM were included in the review,

Table 1 Characteristics of explicit instruments

Characteristic	Instruments (% of the total)
Construct	
Skills	67
Gendered domain	19
Interest	12
Suitability	10
Gender imbalance	10
Attribution	10
Conformance	8
Relevance	8
Domain of interest	
Math/numbers/calculus	67
Science	25
Computing/Programming/ICT	12
Engineering/mechanical	12
Physics	9
Technology/technical	8
Geometry/mental rotation/spatial	7
STEM	7
Chemistry	4
Analytic/reasoning/logic	4
Nature/Geography	2
Astronomy	1
Both STEM and non-STEM domains	38
Non-STEM domains	
Language	13
Native language (e.g., English)	8
Other	8
Reading	7
Arts	7
Liberal arts	4
Humanities	3
Writing	2

Total number of explicit instruments: 91

Total number of studies using an explicit instrument: 104

while those investigating stereotypes' awareness rather than endorsement were excluded. Furthermore, studies using the Draw-a-scientist test were excluded, as these studies have already been summarized in recent reviews (Ferguson and Lezotte 2020; Miller et al. 2018). Finally, studies not in English or Italian were excluded.

One hundred fourteen studies, published from 1993 to the end of 2020 resulted eligible for being included in the review. The instruments used were classified into three macro-categories, i.e., explicit, implicit and indirect (Whitley and Kite 2016). Explicit instruments were, then, evaluated based on the construct measured. In particular, eight types of

Table 2 Characteristics of implicit instruments

Characteristic	Instruments (% of the total)
Type of test	
IAT	52
GNAT	14
Child-IAT	10
IRAP	10
AMP	5
SA-IAT	5
SPF	5
Domain of interest	
Math	48
Science	24
STEM	14
Engineering	10
Space	5
Spatial	5
Both STEM and non-STEM domains	95
Non-STEM domains	
Language	19
Arts	14
Reading	14
Liberal arts	14
Humanities	10
English	10
Other	10
Language arts	5

Total number of implicit instruments: 21

Total number of studies using an implicit instrument: 55

constructs were identified based, in part, on the classification proposed in previous studies (Nurlu 2017; Zitelny et al. 2017).

- *Skills* instruments asking participants to evaluate skills, abilities or brain differences between men and women.
- *Conformance* instruments asking participants to give their opinion about the need for women or men to conform to the opposite gender's behaviour/attitudes in the domain of interest.
- *Gendered domain* instruments generically referring to masculinity or femininity of domains without specifying whether this refers to abilities, interests, or other characteristics.
- *Interest* instruments investigating differences in interest.
- *Relevance* instruments asking participants an opinion on the relevance the domains of interest have for people.
- *Gender imbalance* instruments asking participants to evaluate the representativeness of women and men in areas, occupations, and courses.

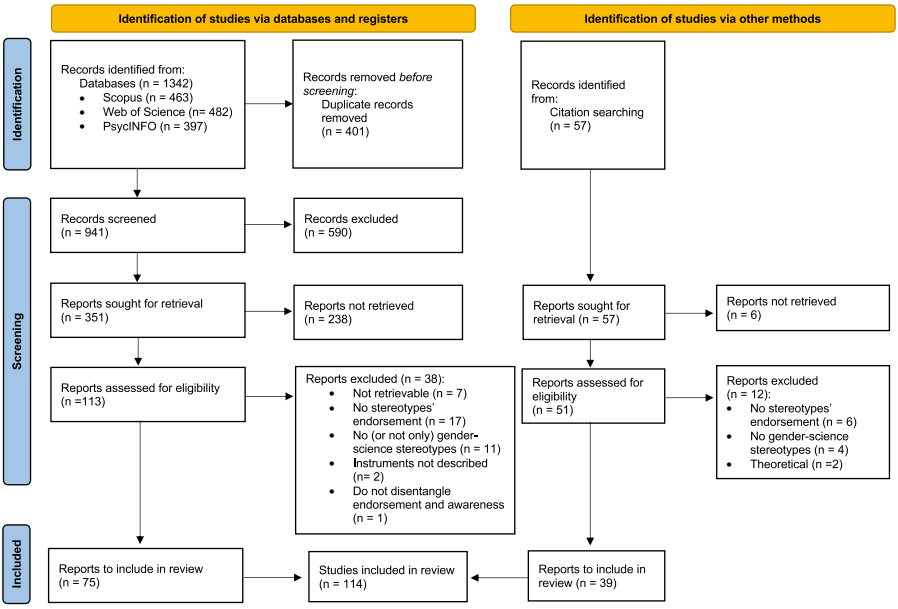

Fig. 1 PRISMA 2020 flow diagram

- *Suitability* instruments asking participants explicitly an opinion on the suitability of people in certain domains.

Other information extracted from the selected studies were the domains of interest (e.g., science, math) and whether and how they measured the association between gender and humanities or language-related skills. Implicit measures were evaluated based on (1) the type of test, and (2) the type of target and categories used (Whitley and Kite 2016), while indirect measures were too few to be further classified.

3 Results

The detailed summary of the instruments included in this review is reported in the Online Appendix. Table A1 lists all explicit instruments and reports for each of them (1) the scale type, (2) the instrument as reported in the questionnaire, (3) the construct measured, (4) the studies in which it was applied and (5) the Cronbach alpha or other reliability indicators if reported. Table A2 lists all implicit instruments and reports (1) the type of test, (2) the target and categories used and (3) the studies in which the instrument was applied. Finally, Table A3 lists the indirect instruments, i.e., how they were measured and the studies in which they were applied.

3.1 Explicit measurements

Most studies included in this report used an explicit measure to assess stereotypes' endorsement. Statements were usually evaluated using Likert scales, the most popular being the 5-point Likert scale.

Table 1 summarizes the characteristics of the explicit measurements. The most frequently investigated construct is that relative to skills, followed by gendered domain, gender imbalance and suitability. However, there are differences even within constructs. These differences regard, in some cases, the content of the items, while in others, the question and the phrasing used for statements (see Table A1 in the Online Appendix).

As regards *skills*, items ask participants to what extent they agree with the belief on the outperformance of men in the STEM field, either directly or indirectly. In other cases, it is asked to rate who is better between men and women, leaving thus the possibility to also detect cases in which the association is even reversed. Finally, in some instruments the cause underlying gender differences is explicitly mentioned, e.g., 'Men are naturally better at advanced math (mechanical things) than women' (Riegle-Crumb and Morton 2017). It is noteworthy that, in one study, researchers showed an explicit interest in distinguishing between descriptive and prescriptive stereotypes (McGuire et al. 2020), the first referring to beliefs on what people do, e.g., 'Who do you think is usually good at […]', the second on what people should do, e.g., 'Who do you think should be good at […]'.

As regards *gendered domain*, questions are similar one to the other and ask participants to rate how much they associate a list of domains with males or females (see, for example, Greenwald et al. 2003; White and White 2006; Young et al. 2013), or whether they agree with statements such as 'Math is rather a typical subject for girls (boys)' (Steffens et al. 2010). Similarly, to assess the opinion on the *representativeness of women and men*, researchers usually ask participants either to estimate the percentage of male and female workers in certain occupations or to provide their agreement with statements on this representativeness, e.g., 'There are more men in science-related jobs' (Breda et al. 2018).

In some cases, participants are asked to compare the *suitability* of women and men to STEM or humanities fields or to give their opinion on the better suitability of men to STEM fields, e.g., 'It is possible that men are better suited to studying at the technical university than women' (Jasko et al. 2019). As for other constructs, questions on gender differences in *interest* ask participants their opinion on the higher interest in STEM of men, e.g., 'Boys (girls) are more interested in careers which require mathematical ability than girls (boys) are' (Nurlu 2017). Similarly, questions on the *relevance* of STEM for men and women, asked, for example, 'It is more important for boys to understand physical science than girls' (Buck et al. 2002). On the contrary, instruments assessing *conformance* are quite different from one another. Ertl et al. (2017) generically asked whether 'Females that are working in the field of STEM have to be like men', Betz and Sekaquaptewa (2012) if 'Do being good at math and being girly go together?', while both Plante et al. (2009) and Nurlu (2017) asked about the association between popularity and abilities in STEM or reading.

In some studies, participants were directly asked their opinion on the potential explanations for gender differences. Contrary to other constructs, in the case of *attribution*, the existence of gender differences is taken for granted. The interest is in verifying whether participants are more likely to attribute the gender gap in STEM to biological

rather than cultural and social factors, e.g., 'Boys (girls) are encouraged more than girls (boys) to choose a career in a math-related area' (Nurlu 2017), 'Males perform better than females in science because of greater natural ability' (Nosek et al. 1998).

Most instruments asked participants for an opinion on 'Math', 25% of them an opinion on 'Science', while in a minority of cases instruments mentioned other STEM-related fields (see Table 1). In some cases (38%), instruments asked an opinion on both STEM and non-STEM fields, the latter being specified in different ways, e.g., 'Language', 'Liberal Arts', 'Humanities'.

3.2 Implicit measurements

While there was far less heterogeneity in the type of implicit measurement generally used to test stereotypes compared to the explicit ones, still there is some variability. The Implicit Association Test (Greenwald et al. 1998) is the most popular measurement of the strength of associations for studying stereotypes. It was designed by Greenwald et al. (1998) to measure individual differences in implicit cognition. This is done by measuring the difference in the time needed to do an association between compatible constructs (e.g., women and humanities, men and STEM) and the time needed to do an association with incompatible constructs (e.g., women and STEM, men and humanities).

However, other tests similar to the IAT were used, i.e., the Implicit Relational Assessment Procedure, IRAP (Barnes-Holmes et al. 2006), the Affect Misattribution Procedure, AMP (Payne et al. 2005), the Go/No-Go Association Task, GNAT (Nosek and Banaji 2001) and the Sorting Paired Feature Task, SPF (Bar-Anan et al. 2009). Contrary to the IAT, the AMP, IRAP and GNAT allow disentangling the two tested associations. Further details on implicit measures can be found in Gawronski and De Houwer (2013) and an application in the case of stereotypes in Whitley and Kite (2016).

Table 2 summarizes the types and characteristics of implicit instruments. More details can be found in the Online Appendix (Table A2).

Similarly to what was observed for explicit instruments but to a lower extent, most instruments used 'Math' as a category while some used 'Science'. However, in the most adopted version of the IAT, *stimuli* were STEM-related majors. As regards non-STEM fields, the choice of the category was more heterogeneous, as *stimuli* referred to either 'Language', 'Arts', 'Reading', or 'Liberal Arts'. Exclusively in one study (Guizzo et al. 2019) only one of the two associations, i.e., gender and space-related concepts, was tested.

3.3 Indirect measurements

Seven studies created and applied indirect instruments (see Table A3 in the Online Appendix). They differ from explicit and implicit measures because there is not an explicit reference to gender. Participants were usually shown two (or more) pictures, one showing a man/boy and the other showing a woman/girl, and asked which one possessed some characteristics, e.g., interest and giftedness in math (Nurnberger et al. 2016). However, two instruments stand out. In Tomasetto et al. (2012), children were told a story about an island where inhabitants would not consider boys and girls equally good in school subjects. At the end of the story, participants were asked whether, in their opinion, the inhabitants of the island considered boys or girls better in math. While, in Ambady et al. (2001), participants were asked to repeat a brief story about a student good in math and the experimenter noted whether they used the pronoun 'he' or 'she' when appointing the student.

3.4 Consequences of instruments' heterogeneity

This review suggests that measurements of gender-science stereotypes show great heterogeneity on a variety of features, i.e., underlying constructs, domains of interest (e.g., science, math, reading, native language), types of scale and number of items, types of stereotypes (descriptive or prescriptive), age of the people on which the belief was asked (children, adolescents, adults), whether the opinion regarded school subjects, majors or occupations, number and type (stereotypical and/or counter-stereotypical) of associations. This variability could have important implications. As mentioned before, some of these implications were already discussed for instruments on gender-roles stereotypes by McHugh and Frieze (1997).

Compared to gender roles, the problem with gender-science beliefs is even more complicated. Indeed, in this case, considering proper scales is impossible because the set of items in the questionnaire was in most studies not chosen following a development process (Kyriazos and Stalikas 2018). Rather, researchers have tended to create ad hoc statements and evaluate them using Likert scales. Assumptions and hypotheses behind items' selection were not usually reported (more details in the Online Appendix). Hence, the problem of validity and reliability of measurements, discussed in McHugh and Frieze (1997), applies even more in this context. However, comparability and reliability of instruments are not the only cons of scale proliferation.

As mentioned before, Zitelny et al. (2017) suggested that the variability in the correlation between explicit measures and the IAT may be due to the use of different self-reported instruments. Therefore, a distinction among constructs of stereotypes and a more specific choice of one over the other may be relevant when both explicit and implicit measurements are used. However, other features of explicit instruments may affect the correlation with the implicit ones. As anticipated, implicit measures' scores, especially for the IAT, are the result of two stereotypical associations, one of men and science, the other of women and humanities. However, only 38% of explicit measurements tested both associations and only in a few cases the final score was computed as the difference between the two (e.g., Liu et al. 2010; Rentas 2015). For instance, Plante and Theoret (2009) created two scales, i.e., the male domain scale and the female domain scale, each including 16 items on abilities, usefulness, attitudes, typicality, effort and support in both math and language. The final score of gender stereotypes was calculated by subtracting the mean scale score of the female domain from the male domain's mean scale score. This distinction is relevant as the reversal of the typical stereotype emerged from the analysis. As suggested by the authors, 'it appears that stereotypes favouring girls in mathematics can emerge when the instrument that is used allows this possibility' (Plante et al. 2009, p. 398). Consequently, implicit and explicit scores reflect two different things.

3.5 Limitations of existing instruments

This review found certain weaknesses caused by the way items were constructed. The first limitation regards the domain associated with gender. As mentioned above, in most cases questions asked an opinion about math. However, women's underrepresentation does not characterize only math-related areas, but more in general the entire scientific field. Indeed, when accounting for the discontinued education and career paths of females in the scientific field, researchers refer to a STEM leaky pipeline, not a math leaky pipeline (Grogan 2019). Among the reviewed studies, while those interested in the association between gender and

occupations usually referred to STEM or science-related careers, those interested in the association between gender and school subjects referred more frequently to math. If the aim is to assess medium- and long-term effects rather than certify the mere existence of stereotypes, extending the domain to include also science could be more appropriate. Note that the gender gap is narrower in math majors compared to other scientific fields of study. In 2018, the percentage of females among U.S. students with a Master's degree was equal to 43% in mathematics and statistics, 38% in physical sciences and science technologies, 32% in computer and information services and 26% in engineering (NCES 2020).

Furthermore, research tended also to disregard the association between gender and reading skills or the corresponding humanities field. As mentioned above, ignoring such an association may be problematic when the implicit association test is combined with explicit measurements. Moreover, the gender gap in humanities-related majors is at least as wide as in STEM majors. In 2018, the percentage of females among U.S. students with a Master's degree in liberal arts and sciences was equal to 39%, in humanities and humanistic studies to 37% (NCES 2020). As suggested by Plante and Theoret (2009) to justify the reversal of beliefs about female math abilities in their sample, initiatives aimed at reducing the underrepresentation of women may contribute to a change in stereotypical beliefs on math, but not on other domains.

Some instruments, especially those on the gendered domain, are based on quite generic statements. In particular, those asking to rate to what extent a certain domain would be feminine or masculine, leave the respondent the freedom to choose which aspects of gender differences the questionnaire is referring to. Femininity may derive from representativeness, which would imply that a domain is feminine when more women than men are working in that area, the opposite for masculine. On the other hand, when considering ability, the domain would be labelled as feminine when the respondent believes that women are better than men in that specific domain. Leaving too much space for interpretations can bias results, as assessing whether the final score would measure the same thing for all individuals in the sample would be impossible. Back to the example above, in one case we would conclude with a descriptive stereotype, whereas in the other case, we would call for a prescriptive stereotype.

The distinction between descriptive and prescriptive stereotypes does not apply only to this case. When posing participants questions about gender imbalance, i.e., different ratios of males and females in majors and occupations, they are required an opinion on what they see rather than what they think, e.g., 'How many men and women usually do this work?'. Similarly, when posing questions on skills, participants are asked their opinion about innate, biological differences in abilities, whereas in other cases, they are only asked about differences in performance. In all these examples, answers could be equally classified as either descriptive or prescriptive behaviours. However, the distinction is not operated leaving authors to consider all these indicators as measurements of the same concept, i.e., gender stereotypes. On the one hand, inferring stereotypes from these different instruments is possible. On the other, as argued above, solid reasons would suggest that making distinctions and refining what is generically called a stereotype is beneficial as descriptive and prescriptive beliefs might have different effects on outcomes.

For instance, in a laboratory experiment on sex discrimination in hiring, Gill (2004) found that descriptive stereotypes did predict gender bias in neither the choice of job applicants nor the evaluation of candidates. On the other hand, results indicated that prescriptive stereotypes fostered a bias among male participants against females enacting a masculine role. Similarly, in their review of the literature on descriptive and prescriptive stereotypes in sex discrimination and sexual harassment, Burgess and Borgida (1999) called for

clear-cut distinctions between these two components as they resulted in different types of sex discrimination. Descriptive stereotypes would lead to an unintentional form of discrimination, which may be modified when information on the inaccuracy of the gender bias is provided. Prescriptive stereotypes lead to a stronger form of discrimination and prejudice, which is not dented by any information. This is not surprising if we link descriptive and prescriptive beliefs to attribution. If we believe the gender gap is due to biological, innate differences, we will be less likely to modify our opinion even when evidence of equality of performance is provided. On the contrary, if other justifications are given for the gender imbalance in the sector, information on inaccuracy is likely to change previous beliefs.

Implications for this psychological process may be quite relevant in the context of gender-science stereotypes, as initiatives aimed at reducing them are based on the exposure to role models (Betz and Sekaquaptewa 2012; Gilbert 2015; Van Camp et al. 2019; Young et al. 2013). Further research would be necessary to understand whether attributions to gender differences in science may affect the efficacy of the exposure to role models, rather than other constructs of stereotypes. Unfortunately, little is known about the effect of these two components in the specific area of gender bias in STEM and humanities. A remarkable exception is McGuire et al. (2020) who collected information on three distinct types of stereotypes, i.e., awareness (descriptive component, e.g., 'who do you think is usually good at…'), endorsement (prescriptive component, e.g., 'who do you think should be good at…') and flexibility (e.g., 'who do you think can be good at…'). However, the authors did not distinguish among the three components when reporting the effect of gender stereotypes on aspirations, performance and other outcomes of interest.

4 Discussion and conclusions

Greater attention has been recently devoted to stereotypes and their influence on gender issues, while these are on top of the political agenda in several countries (Cusack 2013). Academic researchers have, concurrently, conducted empirical studies testing the effect of gender stereotypes on women's engagement in STEM. This has stimulated the creation of scales and other instruments to measure gender stereotypes. However, compared to other fields of gender bias, there is a lack of properly developed scales to assess associations between gender and science/humanities. This led to a proliferation of instruments, which in turn can explain the variability of findings and certain terminological ambiguities.

The current review extended that of Zitelny et al. (2017) on instruments for gender stereotypes in science and summarized implicit, explicit and indirect measures adopted by researchers in 114 articles. Explicit measures were classified based on the underlying construct of stereotypes as follows: attribution to gender differences; conformity to behaviours and attitudes of the prevalent sex in the field; masculinity/femininity of the domain; interest in the subject; representativeness of men and women in the sector; suitability in the domain; performance in the subject; and relevance attributed to the subject by men and women. Research prevalently identified stereotypes with differences in abilities in math/science and reading. However, the instruments differed in several features, such as type of ability, domain investigated, type of scale and nature of stereotypes (descriptive versus prescriptive).

The summary of implicit indicators detected a certain degree of heterogeneity, though less than expected. The most popular test is the Implicit Association Test (Greenwald et al. 1998), yet different versions are used. These versions can be distinguished on the

type of categories and stimuli adopted to construct the test, i.e., words related to majors, occupations, or features of STEM and humanities. As regards indirect measurements, they are mostly adopted when testing stereotypes on children and ask them to associate boys and girls seen in pictures with characteristics related to math/science and language.

It was then discussed certain pitfalls due to the heterogeneity and proliferation of scales. First, the adoption of indicators varying in multiple aspects can eventually invalidate findings' comparison and scale reliability (McHugh and Frieze 1997). This heterogeneity may also explain the variability in the correlation between explicit and implicit measures (Zitelny et al. 2017). While the IAT score is the difference between two associations, i.e., male and science vs female and humanities, explicit measures usually address either one of the two associations or do not construct the final score as a difference. Furthermore, some of the revised instruments may suffer certain potential limitations. In particular, most instruments focus on math, which leaves aside other scientific fields (e.g., science, technology and engineering) and language-related domains, which are affected by stereotypical beliefs as well.

Another limitation regards more specifically the way questions on the masculinity and femininity of science and humanities are posed. Questions of this type are generic and do not specify what it is meant for masculinity or femininity, thereby leaving the interpretation to the respondent. This impairs the final scores of whatever instruments, as they may assume different meanings, e.g., descriptive or prescriptive beliefs. Distinguishing these two beliefs is relevant because they can have different effects on discriminatory behaviours (Burgess and Borgida 1999; Gill 2004).

Finally, it is worth noting that the definitions of stereotypes have changed over time assuming now simpler and less restrictive forms compared to the past. As suggested by Nelson, the latter included also 'inaccuracy, negativity, and overgeneralization. It is unfortunate that we have let those original requirements go — after all, they really are the heart of why we care about the topic at all.' (Nelson 2009, p. 2). Nowadays, we tend to use more 'neutral' definitions, which depict stereotypes as beliefs on groups' characteristics, attributes and behaviours. The existence of multiple definitions and their neutrality are a double-edged sword. On the one hand, they allow us to catch the multifaceted nature of beliefs on gender differences. On the other hand, they led to a proliferation of instruments.

A more accurate choice of instruments in empirical research and refinement of the type of constructs measured by the proposed indicators is required to advance our understanding of these important puzzles. Researchers studying gender stereotypes should prefer using already existing and shared instruments when the aim of their work allows it. Being aware of this, further specification on the subtypes and constructs of gender stereotypes should be given when presenting studies' results. Furthermore, factor analysis should be performed when the chosen items refer to different constructs. Stereotypical beliefs on ability may have different implications than beliefs on the representativeness of women and men in careers on choices, behaviours and attitudes and so biasing our measurements. Eventually, this would contribute to the comprehension of the issue of women in STEM, by facilitating the comparison of similar studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11135-022-01512-8.

Acknowledgements I thank Flaminio Squazzoni (University of Milan) for comments that greatly improved the manuscript. I am thankful to the editor and the anonymous reviewers for their remarkable suggestions.

Funding Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement. No funding was received for conducting this study.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Ambady, N., Shih, M., Kim, A., Pittinsky, T.L.: Stereotype susceptibility in children: effects of identity activation on quantitative performance. Psychol. Sci. **12**(5), 385–390 (2001). https://doi.org/10.1111/1467-9280.00371

Bar-Anan, Y., Nosek, B., Vianello, M.: The sorting paired features task: a measure of association strengths. Exp. Psychol. **56**(5), 329–343 (2009). https://doi.org/10.1027/1618-3169.56.5.329

Barnes-Holmes, D., Barnes-Holmes, Y., Hayden, E., Milne, R., Stewart, I.B.: Do you really know what you Believe? developing the implicit relational assessment procedure (IRAP) as a direct measure of implicit beliefs. Irish Psychol. **32**(7), 169–177 (2006)

Beere, C.A.: Gender Roles: A Handbook of Tests and Measures. Greenwood Press. (1990). https://doi.org/10.1177/027046769301300329

Begeny, C.T., Ryan, M.K., Moss-Racusin, C.A., Ravetz, G.: In some professions, women have become well represented, yet gender bias persists: perpetuated by those who think it is not happening. Sci. Adv. **6**(26), eaba7814 (2020). https://doi.org/10.1126/sciadv.aba7814

Betz, D.E., Sekaquaptewa, D.: My fair physicist? feminine math and science role models demotivate young girls. Soc. Psychol. Personal. Sci. **3**(6), 738–746 (2012). https://doi.org/10.1177/1948550612440735

Breda, T., Grenet, J., Monnet, M., Effenterre, C. van.: Can Female Role Models Reduce the Gender Gap in Science? Evidence from Classroom Interventions in French High Schools (Halshs-01713068; PSE Working Papers). HAL. (2018). https://ideas.repec.org/p/hal/psewpa/halshs-01713068.html

Buck, G.A., Leslie-Pelecky, D., Kirby, S.K.: Bringing female scientists into the elementary classroom: confronting the strength of elementary students' stereotypical images of scientists. J. Elem. Sci. Educ. **14**(2), 1–9 (2002)

Burgess, D., Borgida, E.: Who women are, who women should be: descriptive and prescriptive gender stereotyping in sex discrimination. Psychol. Public Policy Law **5**(3), 665–692 (1999). https://doi.org/10.1037/1076-8971.5.3.665

Cundiff, J.L., Vescio, T.K., Loken, E., Lo, L.: Do gender–science stereotypes predict science identification and science career aspirations among undergraduate science majors? Soc. Psychol. Educ. **16**(4), 541–554 (2013). https://doi.org/10.1007/s11218-013-9232-8

Cusack, S.: Gender Stereotyping as a Human Rights Violation. Office of the High Commissioner for Human Rights. (2013). https://www.esem.org.mk/pdf/Najznachajni%20vesti/2014/3/Cusack.pdf

Dasgupta, N., Asgari, S.: Seeing is believing: exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. J. Exp. Soc. Psychol. **40**(5), 642–658 (2004). https://doi.org/10.1016/j.jesp.2004.02.003

Dwyer, C.A.: Influence of children's sex role standards on reading and arithmetic achievement. J. Educ. Psychol. **66**(6), 811–816 (1974). https://doi.org/10.1037/h0021522

Eagly, A. H., Wood, W.: Social role theory. In: Handbook of Theories of Social Psychology, Vol. 2, pp. 458–476. SAGE Publications Ltd. (2012). https://doi.org/10.4135/9781446249222

Ellemers, N.: Gender stereotypes. Annu. Rev. Psychol. **69**(1), 275–298 (2018). https://doi.org/10.1146/annurev-psych-122216-011719

Ertl, B., Luttenberger, S., Paechter, M.: The Impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females. Front. Psychol. **8**, 703 (2017). https://doi.org/10.3389/fpsyg.2017.00703

Farrell, L., Nearchou, F., McHugh, L.: Examining the effectiveness of brief interventions to strengthen a positive implicit relation between women and STEM across two timepoints. Soc. Psychol. Educ. **23**(5), 1203–1231 (2020). https://doi.org/10.1007/s11218-020-09576-w

Fennema, E., Sherman, J.A.: Fennema-Sherman mathematics attitudes scales: instruments designed to measure attitudes toward the learning of mathematics by females and males. J. Res. Math. Educ. **7**(5), 324–326 (1976)

Ferguson, S., Lezotte, S.: Exploring the state of science stereotypes: systematic review and meta-analysis of the draw-a-scientist checklist. Sch. Sci. Math. **120**(1), 55–65 (2020). https://doi.org/10.1111/ssm.12382

Gawronski, B., De Houwer, J.: Implicit Measures in social and personality psychology. In: Reis, H.T., Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology*,2nd ed., pp. 283–310. Cambridge University Press. (2013). https://doi.org/10.1017/CBO9780511996481.016

Gilbert, P.N., O'Brien, L.T., Garcia, D.M., Marx, D.M.: Not the sum of its parts: decomposing implicit academic stereotypes to understand sense of fit in math and english. Sex Roles **72**(1–2), 25–39 (2015). https://doi.org/10.1007/s11199-014-0428-y

Gilbert, P.N.: The role of role models: How does identification with STEM role models impact women's implicit STEM stereotypes and STEM outcomes? [Ph.D. Dissertation, Tulane University]. (2015) https://digitallibrary.tulane.edu/islandora/object/tulane%3A27945

Gill, M.J.: When information does not deter stereotyping: prescriptive stereotyping can foster bias under conditions that deter descriptive stereotyping. J. Exp. Soc. Psychol. **40**(5), 619–632 (2004). https://doi.org/10.1016/j.jesp.2003.12.001

Glick, P., Fiske, S.T.: The ambivalent sexism inventory: differentiating hostile and benevolent sexism. J. Pers. Soc. Psychol. **70**(3), 491–512 (1996). https://doi.org/10.1037/0022-3514.70.3.491

Greenwald, A.G., Farnham, S.D.: Using the implicit association test to measure self-esteem and self-concept. J. Pers. Soc. Psychol. **79**(6), 1022–1038 (2000). https://doi.org/10.1037/0022-3514.79.6.1022

Greenwald, A.G., McGhee, D.E., Schwartz, J.L.K.: Measuring individual differences in implicit cognition: the implicit association test. J. Pers. Soc. Psychol. **74**(6), 1464–1480 (1998). https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A.G., Nosek, B., Banaji, M.R.: Understanding and using the implicit association test: I. An improved scoring algorithm. J. Personal. Soc. Psychol. **85**(2), 197–216 (2003). https://doi.org/10.1037/0022-3514.85.2.197

Grogan, K.E.: How the entire scientific community can confront gender bias in the workplace. Nat. Ecol. Evol. **3**(1), 3–6 (2019). https://doi.org/10.1038/s41559-018-0747-4

Grow, A., Takács, K., Pál, J.: Status characteristics and ability attributions in hungarian school classes: an exponential random graph approach. Soc. Psychol. Q. **79**(2), 156–167 (2016). https://doi.org/10.1177/0190272516643052

Guizzo, F., Moe, A., Cadinu, M., Bertolli, C.: The role of implicit gender spatial stereotyping in mental rotation performance. Acta Physiol. (oxf) **194**, 63–68 (2019). https://doi.org/10.1016/j.actpsy.2019.01.013

Hentschel, T., Heilman, M.E., Peus, C.V.: The Multiple dimensions of gender stereotypes: a current look at men's and women's characterizations of others and themselves. Front. Psychol. **10**, 11 (2019). https://doi.org/10.3389/fpsyg.2019.00011

Hilton, J.L., von Hippel, W.: STEREOTYPES. Annu. Rev. Psychol. **47**(1), 237–271 (1996). https://doi.org/10.1146/annurev.psych.47.1.237

Jackson, S.M., Hillard, A.L., Schneider, T.R.: Using implicit bias training to improve attitudes toward women in STEM. Soc. Psychol. Educ. **17**(3), 419–438 (2014). https://doi.org/10.1007/s11218-014-9259-5

Jasko, K., Dukala, K., Szastok, M.: Focusing on gender similarities increases female students' motivation to participate in STEM. J. Appl. Soc. Psychol. **49**(8), 473–487 (2019). https://doi.org/10.1111/jasp.12598

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

Jost, J.T., Kay, A.C.: Exposure to benevolent sexism and complementary gender stereotypes: consequences for specific and diffuse forms of system justification. J. Pers. Soc. Psychol. **88**(3), 498–509 (2005). https://doi.org/10.1037/0022-3514.88.3.498

Kersey, A.J., Csumitta, K.D., Cantlon, J.F.: Gender similarities in the brain during mathematics development. Npj Sci. Learn. **4**(1), 19 (2019). https://doi.org/10.1038/s41539-019-0057-x

Kiefer, A.K., Sekaquaptewa, D.: Implicit stereotypes, gender identification, and math-related outcomes: a prospective study of female college students. Psychol. Sci. **18**(1), 13–18 (2007a). https://doi.org/10.1111/j.1467-9280.2007.01841.x

Kiefer, A.K., Sekaquaptewa, D.: Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. J. Exp. Soc. Psychol. **43**(5), 825–832 (2007b). https://doi.org/10.1016/j.jesp.2006.08.004

Kite, M.E., Deaux, K., Haines, E.L.: Gender stereotypes. In Psychology of Women: A Handbook of Issues and Theories, 2nd ed., pp. 205–236. (2008). Praeger Publishers/Greenwood Publishing Group

Kyriazos, T.A., Stalikas, A.: Applied psychometrics: the steps of scale development and standardization process. Psychology **09**(11), 2531–2560 (2018). https://doi.org/10.4236/psych.2018.911145

Lane, K.A., Goh, J.X., Driver-Linn, E.: Implicit science stereotypes mediate the relationship between gender and academic participation. Sex Roles **66**(3–4), 220–234 (2012). https://doi.org/10.1007/s11199-011-0036-z

Leder, G., Forgasz, H.: Two New Instruments to Probe Attitudes About Gender and Mathematics (p. 29). La Trobe University. (2002). https://www.academia.edu/20451640/Two_New_Instruments_To_Probe_Attitudes_about_Gender_and_Mathematics

Lippmann, W.: Public Opinion. Harcourt, Brace & Co. (1922)

Liu, M., Hu, W., Jiannong, S., Adey, P.: Gender stereotyping and affective attitudes towards science in Chinese secondary school students. Int. J. Sci. Educ. **32**(3), 379–395 (2010). https://doi.org/10.1080/09500690802595847

Marx, D.M., Ko, S.J.: Prejudice, discrimination, and stereotypes (Racial Bias). In: Encyclopedia of Human Behavior, pp. 160–166. Elsevier, (2012). https://doi.org/10.1016/B978-0-12-375000-6.00388-8

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

McGuire, L., Mulvey, K.L., Goff, E., Irvin, M.J., Winterbottom, M., Fields, G.E., Hartstone-Rose, A., Rutland, A.: STEM gender stereotypes from early childhood through adolescence at informal science centers. J. Appl. Dev. Psychol. **67**, 101109 (2020). https://doi.org/10.1016/j.appdev.2020.101109

McHugh, M.C., Frieze, I.H.: The Measurement of gender-role attitudes: a review and commentary. Psychol. Women Q. **21**(1), 1–16 (1997). https://doi.org/10.1111/j.1471-6402.1997.tb00097.x

McIntyre, R.B., Paulson, R.M., Lord, C.G.: Alleviating women's mathematics stereotype threat through salience of group achievements. J. Exp. Soc. Psychol. **39**(1), 83–90 (2003). https://doi.org/10.1016/S0022-1031(02)00513-9

Miller, D.I., Nolla, K.M., Eagly, A.H., Uttal, D.H.: The development of children's gender-science stereotypes: a meta-analysis of 5 decades of U.S. draw-a-scientist studies. Child Dev. **89**(6), 1943–1955 (2018). https://doi.org/10.1111/cdev.13039

NCES. (2020). Table 318.30: Bachelor's, Master's, and Doctor's Degrees Conferred by Postsecondary Institutions, By Sex of Student and Discipline Division: 2018–19. National Center for Education Statistics. https://nces.ed.gov/programs/digest/d20/tables/dt20_318.30.asp

Nelson, T.D. (ed.): Handbook of prejudice, stereotyping, and discrimination. Psychology Press (2009)

Nosek, B., Banaji, M.R.: The Go/No-Go association task. Soc. Cogn. **19**(6), 625–666 (2001). https://doi.org/10.1521/soco.19.6.625.20886

Nosek, B., Smyth, F.L.: Implicit social cognitions predict sex differences in math engagement and achievement. Am. Educ. Res. J. **48**(5), 1125–1156 (2011). https://doi.org/10.3102/0002831211410683

Nosek, B., Banaji, M.R., Greenwald, A.G.: Project Implicit. (1998). https://implicit.harvard.edu/implicit/index.jsp

Nurlu, Ö.: Developing a teachers gender stereotype scale toward mathematics. Int. Electron. J. Elem. Educ. **10**(2), 287–299 (2017). https://doi.org/10.26822/iejee.2017236124

Nurnberger, M., Nerb, J., Schmitz, F., Keller, J., Sutterlin, S.: Implicit gender stereotypes and essentialist beliefs predict preservice teachers' tracking recommendations. J. Exp. Educ. **84**(1), 152–174 (2016). https://doi.org/10.1080/00220973.2015.1027807

Payne, B.K., Cheng, C.M., Govorun, O., Stewart, B.D.: An inkblot for attitudes: affect misattribution as implicit measurement. J. Pers. Soc. Psychol. **89**(3), 277–293 (2005). https://doi.org/10.1037/0022-3514.89.3.277

Plante, I., Théorêt, M., Favreau, O.E.: Student gender stereotypes: contrasting the perceived maleness and femaleness of mathematics and language. Educ. Psychol. **29**(4), 385–405 (2009). https://doi.org/10.1080/01443410902971500

Rentas, C.A.: The effects of a perceived causal relationship on the strength of stereotypes. Diss. Abstr. Int. Sect. A: Human. Soc. Sci. 76(6-A(E)) (2015)

Retelsdorf, J., Schwartz, K., Asbrock, F.: "Michael can't read!" Teachers' gender stereotypes and boys' reading self-concept. J. Educ. Psychol. **107**(1), 186–194 (2015). https://doi.org/10.1037/a0037107
Reuben, E., Sapienza, P., Zingales, L.: How stereotypes impair women's careers in science. Proc. Natl. Acad. Sci. **111**(12), 4403–4408 (2014). https://doi.org/10.1073/pnas.1314788111
Riegle-Crumb, C., Morton, K.: Gendered expectations: examining how peers shape female students' intent to pursue STEM fields. Front. Psychol. (2017). https://doi.org/10.3389/fpsyg.2017.00329
Schneider, D.J.: The psychology of stereotyping (Paperback ed). Guilford Press (2005)
Sczesny, S., Nater, C., Eagly, A.H.: Agency and communion: their implications for gender stereotypes and gender identities. Agency Commun. Soc. Psychol. (2018). https://doi.org/10.4324/9780203703663
Six, B., Eckes, T.: A closer look at the complex structure of gender stereotypes. Sex Roles **24**(1), 57–71 (1991). https://doi.org/10.1007/BF00288703
Smiler, A.P.: Thirty years after the discovery of gender: psychological concepts and measures of masculinity. Sex Roles **50**(1/2), 15–26 (2004). https://doi.org/10.1023/B:SERS.0000011069.02279.4c
Spence, J.T., Helmreich, R.: The attitudes toward women scale: an objective instrument to measure attitudes toward the rights and roles of women in contemporary society. Catalog Sel. Doc. Psychol. 2(66) (1972)
Steffens, M.C., Jelenec, P., Noack, P.: On the leaky math pipeline: comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. J. Educ. Psychol. **102**(4), 947–963 (2010). https://doi.org/10.1037/a0019920
Tomasetto, C., Galdi, S., Cadinu, M.: Quando l'implicito precede l'esplicito: Gli stereotipi di genere sulla matematica in bambine e bambini di 6 anni. Psicol. Soc. **2**, 169–186 (2012). https://doi.org/10.1482/37693
Van Camp, A.R., Gilbert, P.N., O'Brien, L.T.: Testing the effects of a role model intervention on women's STEM outcomes. Soc. Psychol. Educ. **22**(3), 649–671 (2019). https://doi.org/10.1007/s11218-019-09498-2
White, M.J., White, G.B.: Implicit and explicit occupational gender stereotypes. Sex Roles **55**(3–4), 259–266 (2006). https://doi.org/10.1007/s11199-006-9078-z
Whitley, B.E., Kite, M.E.: Psychology of Prejudice and Discrimination (Third Edition). Routledge/Taylor & Francis Group (2016)
Young, D.M., Rudman, L.A., Buettner, H.M., McLean, M.C.: The influence of female role models on women's implicit science cognitions. Psychol. Women Q. **37**(3), 283–292 (2013). https://doi.org/10.1177/0361684313482109
Zitelny, H., Shalom, M., Bar-Anan, Y.: What is the implicit gender-science stereotype? exploring correlations between the gender-science IAT and self-report measures. Soc. Psychol. Personal. Sci. **8**(7), 719–735 (2017). https://doi.org/10.1177/1948550616683017