

Tine and Gotlieb (2013)

EPPI-Centre (2003) & Critical Appraisal Skills Programme (2018)

If the study has a broad focus and this data extraction focuses on just one component of the study, please specify this here

- ☒ Not applicable (whole study is focus of data extraction)
- ☐ Specific focus of this data extraction (please specify)

Study aim(s) and rationale

Was the study informed by, or linked to, an existing body of empirical and/or theoretical research?

Please write in authors' declaration if there is one. Elaborate if necessary, but indicate which aspects are reviewers' interpretation.

- ☒ Explicitly stated (please specify)
- ☐ Implicit (please specify)
- ☐ Not stated/unclear (please specify)
 - Stereotype threat
 - Achievement gaps between different groups
 - Single minority stereotype threat effects on maths performance
 - Multiple minority stereotype threat effects on maths performance
 - Single minority stereotype threat effects on working memory
 - Multiple minority stereotype threat effects on working memory.

Do authors report how the study was funded?

- ☒ Explicitly stated (please specify)
- ☐ Implicit (please specify)
- ☐ Not stated/unclear (please specify)

This research was supported by a Kaminsky Family Fund award and the Paul K. Richter and Evalyn E. Cook Richter Memorial Fund.

Study research question(s) and its policy or practice focus

What is/are the topic focus/foci of the study?

- Stereotype threat effects on working memory and effort will be examined in the current study, in addition to stereotype threat effects on mathematics performance.
- The first goal is to document single minority gender-, race-, and income-based stereotype threat effects on maths performance and working memory function.
- This goal affords us the opportunity to determine the relative impact of gender-, race-, and income-based standard stereotype threat effects, which has not previously been done
- The second goal of the present study is to determine whether there is a multiple minority stereotype threat effect on maths performance and/or working memory function.

What is/are the population focus/foci of the study?

- individuals under stereotype threat (gender, race and/or income)

What is the relevant age group?

- ☐ Not applicable (focus not learners)
- ☐ 0 - 4
- ☐ 5 - 10
- ☐ 11 - 16
- ☐ 17 - 20
- ☐ 21 and over
- ☒ Not stated/unclear

What is the sex of the population focus/foci?

- ☐ Not applicable (focus not learners)
- ☐ Female only
- ☐ Male only
- ☒ Mixed sex

☐ Not stated/unclear

What is/are the educational setting(s) of the study?

- ☐ Community centre
- ☐ Correctional institution
- ☐ Government department
- ☐ Higher education institution
- ☐ Home
- ☐ Independent school
- ☐ Local education authority
- ☐ Nursery school
- ☐ Other early years setting
- ☐ Post-compulsory education institution
- ☐ Primary school
- ☐ Residential school
- ☐ Secondary school
- ☐ Special needs school
- ☐ Workplace
- ☐ Other educational setting

In Which country or countries was the study carried out?

- ☒ Explicitly stated (please specify)
- ☐ Not stated/unclear (please specify)
 - United states of America

Please describe in more detail the specific phenomena, factors, services, or interventions with which the study is concerned

What are the study research questions and/or hypotheses?

Research questions or hypotheses operationalise the aims of the study. Please write in authors' description if there is one. Elaborate if necessary, but indicate which aspects are reviewers' interpretation.

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

- It is expected that the effect of gender-based stereotype threat will be weaker than the effect of race- or income-based stereotype threat.
- We hypothesize that income- and race-based stereotype threat effects may be more extreme than gender-based effects, as stigmatized income- and race-groups are smaller in size than the stigmatized gender group.
- It is hypothesized that there will be an additive multiple minority stereotype threat effect on both outcomes (maths performance and/or working memory function).
- That is, it is hypothesized that individuals with multiple stigmatized aspects of identity will experience more severe stereotype threat effects when primed than those who have fewer stigmatized aspects of identity.
- the hypotheses regarding the impact on effort remains open, as the literature reports contradictory findings.

Methods - Design

Which variables or concepts, if any, does the study aim to measure or examine?

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

- stigmatized aspect of identity (gender, race and/or income)
- effects on maths performance
- effects on working memory

Study timing

Please indicate all that apply and give further details where possible.

If the study examines one or more samples, but each at only one point in time it is cross-sectional.

If the study examines the same samples, but as they have changed over time, it is retrospective, provided that the interest is in starting at one timepoint and looking backwards over time.

If the study examines the same samples as they have changed over time and if data are collected forward over time, it is prospective provided that the interest is in starting at one timepoint and looking forward in time.

- ☒ Cross-sectional
- ☐ Retrospective
- ☐ Prospective
- ☐ Not stated/unclear (please specify)

If the study is an evaluation, when were measurements of the variable(s) used for outcome made, in relation to the intervention?

If at least one of the outcome variables is measured both before and after the intervention, please use the before and after category.

- ☐ Not applicable (not an evaluation)
- ☒ Before and after
- ☐ Only after
- ☐ Other (please specify)
- ☐ Not stated/unclear (please specify)

Methods - Groups

If comparisons are being made between two or more groups, please specify the basis of any divisions made for making these comparisons.

Please give further details where possible.

- ☐ Not applicable (not more than one group)
- ☐ Prospective allocation into more than one group (e.g. allocation to different interventions, or allocation to intervention and control groups)

☒ No prospective allocation but use of pre-existing differences to create comparison groups (e.g. receiving different interventions, or characterised by different levels of a variable such as social class)

☐ Other (please specify)

☐ Not stated/unclear (please specify)

How do the groups differ?

☐ Not applicable (not more than one group)

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

- 2 (gender: male vs female) x 2 (race: White vs race minority) x 3 (income: high vs. middle vs. low)
- based on these, the researchers created 4 groups, each with a different combination of “NASI” scores, Number of Stigmatized Aspects of Identity.
- The number of categories (2 x 2 x 3) in which each participant was stigmatized was summed to create a measure of their total NASI, ranging from zero to three.
- the groups differ in the amount of stigmatized aspects of identity they have.

Number of groups

For instance, in studies in which comparisons are made between groups, this may be the number of groups into which the dataset is divided for analysis (e.g. social class, or form size), or the number of groups allocated to, or receiving, an intervention.

☐ Not applicable (not more than one group)

☐ One

☐ Two

☐ Three

☒ Four or more (please specify)

☐ Other/unclear (please specify)

- see above

Was the assignment of participants to interventions randomised?

☐ Not applicable (not more than one group)

☒ Not applicable (no prospective allocation)

☐ Random

☐ Quasi-random

☐ Non-random

☐ Not stated/unclear (please specify)

Where there was prospective allocation to more than one group, was the allocation sequence concealed from participants and those enrolling them until after enrolment?

Bias can be introduced, consciously or otherwise, if the allocation of pupils or classes or schools to a programme or intervention is made in the knowledge of key characteristics of those allocated. For example: children with more serious reading difficulty might be seen as in greater need and might be more likely to be allocated to the 'new' programme, or the opposite might happen. Either would introduce bias.

☐ Not applicable (not more than one group)

☒ Not applicable (no prospective allocation)

☐ Yes (please specify)

☐ No (please specify)

☐ Not stated/unclear (please specify)

Apart from the experimental intervention, did each study group receive the same level of care (that is, were they treated equally)?

☒ Yes

☐ No

☐ Can't tell

Study design summary

In addition to answering the questions in this section, describe the study design in your own words. You may want to draw upon and elaborate the answers you have already

given.

- All participants were exposed to the same testing condition
1. Came into the lab, White female experimenter explained procedure
 2. They were to take two tests, each divided into two sections to understand underlying cognitive factors involved in maths problem solving
 3. 10 minutes to work on the maths pre-test
 4. then working memory pre-test
 5. priming with gender-, race-, and income-based stereotypes
 6. maths post-test
 7. primed again
 8. working memory post-test
 9. experiment experience and demographics survey
 10. thanked, debriefed, and compensated for their time.

Methods - Sampling strategy

Are the authors trying to produce findings that are representative of a given population?

Please write in authors' description. If authors do not specify please indicate reviewers' interpretation.

- ☐ Explicitly stated (please specify)
- ☒ Implicit (please specify)
- ☐ Not stated/unclear (please specify)
- individuals that belong to a stigmatized group (gender, race and/or income)

Which methods does the study use to identify people or groups of people to sample from and what is the sampling frame?

e.g. telephone directory, electoral register, postcode, school listing, etc. There may be two stages – e.g. first sampling schools and then classes or pupils within them.

- ☐ Not applicable (please specify)
- ☐ Explicitly stated (please specify)
- ☒ Implicit (please specify)
- ☐ Not stated/unclear (please specify)
- undergraduate college students

Which methods does the study use to select people or groups of people (from the sampling frame)?

e.g. selecting people at random, systematically - selecting for example every 5th person, purposively in order to reach a quota for a given characteristic.

- ☐ Not applicable (no sampling frame)
- ☐ Explicitly stated (please specify)
- ☐ Implicit (please specify)
- ☒ Not stated/unclear (please specify)

Planned sample size

If more than one group please give details for each group separately.

- ☐ Not applicable (please specify)
- ☐ Explicitly stated (please specify)
- ☒ Not stated/unclear (please specify)

Methods - Recruitment and consent

Which methods are used to recruit people into the study?

e.g. letters of invitation, telephone contact, face-to-face contact.

- ☐ Not applicable (please specify)
- ☐ Explicitly stated (please specify)
- ☒ Implicit (please specify)
- ☐ Not stated/unclear (please specify)
- University

Were any incentives provided to recruit people into the study?

- ☐ Not applicable (please specify)
- ☒ Explicitly stated (please specify)
- ☐ Not stated/unclear (please specify)
- “they were compensated for their time”

Was consent sought?

Please comment on the quality of consent if relevant.

- ☐ Not applicable (please specify)
- ☐ Participant consent sought
- ☐ Parental consent sought
- ☐ Other consent sought
- ☐ Consent not sought
- ☒ Not stated/unclear (please specify)

Are there any other details relevant to recruitment and consent?☐ No☒ Yes (please specify)

- The mean quantitative SAT maths reasoning test score was 694 (SD = 90.67), and no participants scored below 450 out of 800

Methods - Actual sample***What was the total number of participants in the study (the actual sample)?****If more than one group is being compared please give numbers for each group.*☐ Not applicable (e.g. study of policies, documents, etc)☒ Explicitly stated (please specify)☐ Implicit (please specify)☐ Not stated/unclear (please specify)

- Seventy-one (N = 71) undergraduate college students between ages 18 and 26 (M = 19.54 years, SD = 1.602) participated.
- Twenty-five (n = 25) participants self identified as male. Forty-six (n = 46) self identified as female and were considered to have a gender-based stigmatized aspect of identity (SAI).
- In total there were 24 racial/ethnic minority (RM) participants; 17 of these identified as racially Black or African American. Seven (n = 7) participants identified as ethnically Hispanic or Latino(a) (and did not identify as racially White). These 24 RM participants were considered to have a race-based SAI. The other 47 (n = 47) participants identified as White.
- Participants were categorized as low-, middle-, or high-income based on an updated eight-income-bracket scheme used by Harrison et al. (2006). The 15 (n = 15) participants who were low-income were categorized as having an income-based SAI, consistent with previous work that found stereotype threat effects on an academic task for low, but not for middle or high-income participants (Harrison et al., 2006).

What is the proportion of those selected for the study who actually participated in the study?*Please specify numbers and percentages if possible.*☐ Not applicable (e.g. study of policies, documents, etc)☐ Explicitly stated (please specify)☐ Implicit (please specify)☒ Not stated/unclear (please specify)

Which country/countries are the individuals in the actual sample from?

If UK, please distinguish between England, Scotland, N. Ireland, and Wales if possible. If from different countries, please give numbers for each. If more than one group is being compared, please describe for each group.

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☐ Explicitly stated (please specify)
- ☐ Implicit (please specify)
- ☒ Not stated/unclear (please specify)

What ages are covered by the actual sample?

Please give the numbers of the sample that fall within each of the given categories. If necessary, refer to a page number in the report (e.g. for a useful table). If more than one group is being compared, please describe for each group. If follow-up study, age at entry to the study.

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☐ 0 to 4
- ☐ 5 to 10
- ☐ 11 to 16
- ☐ 17 to 20
- ☐ 21 and over
- ☒ Not stated/unclear (please specify)

What is the socio-economic status of the individuals within the actual sample?

If more than one group is being compared, please describe for each group.

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☒ Explicitly stated (please specify)
- ☐ Implicit (please specify)
- ☐ Not stated/unclear (please specify)
- 15 participants were considered low-income.

What is the ethnicity of the individuals within the actual sample?

If more than one group is being compared, please describe for each group.

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☐ Explicitly stated (please specify)
- ☐ Implicit (please specify)
- ☐ Not stated/unclear (please specify)

What is known about the special educational needs of individuals within the actual sample?

e.g. specific learning, physical, emotional, behavioural, intellectual difficulties.

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☐ Explicitly stated (please specify)
- ☐ Implicit (please specify)
- ☒ Not stated/unclear (please specify)

Is there any other useful information about the study participants?

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☐ Explicitly stated (please specify no/s.)
- ☒ Implicit (please specify)
- ☐ Not stated/unclear (please specify)
- Two different power analyses programs were used to determine an appropriate sample size (Power in Two-Level Designs and Optimal Design), each with unique flexibilities (Bosker et al. 2003; Raudenbush et al. 2011, respectively).
- Effect sizes and standard deviations were based on prior research. Together, these two programs allowed consideration for the full range of planned analyses.

How representative was the achieved sample (as recruited at the start of the study) in relation to the aims of the sampling frame?

Please specify basis for your decision.

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☐ Not applicable (no sampling frame)
- ☐ High (please specify)
- ☐ Medium (please specify)
- ☐ Low (please specify)
- ☒ Unclear (please specify)

If the study involves studying samples prospectively over time, what proportion of the sample dropped out over the course of the study?

If the study involves more than one group, please give drop-out rates for each group separately. If necessary, refer to a page number in the report (e.g. for a useful table).

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☒ Not applicable (not following samples prospectively over time)
- ☐ Explicitly stated (please specify)
- ☐ Implicit (please specify)
- ☐ Not stated/unclear

For studies that involve following samples prospectively over time, do the authors provide any information on whether and/or how those who dropped out of the study differ from those who remained in the study?

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☒ Not applicable (not following samples prospectively over time)

- ☐ Not applicable (no drop outs)
- ☐ Yes (please specify)
- ☐ No

If the study involves following samples prospectively over time, do authors provide baseline values of key variables such as those being used as outcomes and relevant socio-demographic variables?

- ☐ Not applicable (e.g. study of policies, documents, etc)
- ☒ Not applicable (not following samples prospectively over time)
- ☐ Yes (please specify)
- ☐ No

Methods - Data collection

Please describe the main types of data collected and specify if they were used (a) to define the sample; (b) to measure aspects of the sample as findings of the study?

- ☐ Details
 - SAT maths reasoning scores -> a
 - demographics -> a and b
 - NASI scores -> b
 - maths performance pre- and post-test -> b
 - working memory pre- and post-test -> b
 - experience, effort survey -> b

Which methods were used to collect the data?

Please indicate all that apply and give further detail where possible.

- ☐ Curriculum-based assessment
- ☐ Focus group
- ☐ Group interview
- ☐ One to one interview (face to face or by phone)
- ☐ Observation
- ☐ Self-completion questionnaire
- ☐ Self-completion report or diary
- ☐ Exams
- ☐ Clinical test
- ☐ Practical test
- ☐ Psychological test
- ☐ Hypothetical scenario including vignettes
- ☐ School/college records (e.g. attendance records etc)
- ☐ Secondary data such as publicly available statistics

- ☐ Other documentation
- ☐ Not stated/unclear (please specify)

Details of data collection methods or tool(s).

Please provide details including names for all tools used to collect data and examples of any questions/items given. Also please state whether source is cited in the report.

- ☒ Explicitly stated (please specify)
- ☐ Implicit (please specify)
- ☐ Not stated/unclear (please specify)
- aspect of identity: race -> self-reported survey
- Aspect of identity: income -> adjusted to reflect inflation, based on an updated eight-income-bracket scheme used by Harrison et al. (2006). Participants from a family whose total annual income was < \$45,000 a year were categorized as low-income (n = 15), participants from a family with an annual income between \$45,000 and \$84,999 a year were categorized as middle-income (n = 18), and participants from a family earning over \$85,000 a year were categorized as high-income (n = 38).
- Stereotype threat manipulation: priming script based on previous stereotype threat prime scripts (see Beilock et al. 2007; Rydell et al. 2009).
- Maths pre- and post-test: word problems involving advanced algebra. Test, was originally constructed by Schmader (2002) and used by Rydell and Boucher (2010) and Rydell et al. (2009), was divided into two sections of equal difficulty.
- Working memory pre- and post-test: adapted from Alloway (2007) Automated Working Memory Assessment (AWMA) backward digit string lists and forward digit string lists.
- Experiment experience, effort, and demographics survey: 21-item survey.

Who collected the data?

Please indicate all that apply and give further detail where possible.

- ☐ Researcher
- ☐ Head teacher/Senior management
- ☐ Teaching or other staff
- ☐ Parents
- ☐ Pupils/students
- ☐ Governors
- ☐ LEA/Government officials
- ☐ Other education practitioner
- ☐ Other (please specify)
- ☐ Not stated/unclear

Do the authors describe any ways they addressed the reliability of their data collection tools/methods?

e.g. test-retest methods (Where more than one tool was employed please provide details for each.)

☐ Details

Do the authors describe any ways they have addressed the validity of their data collection tools/methods?

e.g. mention previous validation of tools, published version of tools, involvement of target population in development of tools. (Where more than one tool was employed please provide details for each.)

☐ Details

Was there concealment of study allocation or other key factors from those carrying out measurement of outcome – if relevant?

Not applicable – e.g. analysis of existing data, qualitative study. No – e.g. assessment of reading progress for dyslexic pupils done by teacher who provided intervention. Yes – e.g. researcher assessing pupil knowledge of drugs - unaware of pupil allocation.

☐ Not applicable (please say why)

☐ Yes (please specify)

☐ No (please specify)

Where were the data collected?

e.g. school, home.

☐ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Unclear/not stated (please specify)

Are there other important features of data collection?

e.g. use of video or audio tape; ethical issues such as confidentiality etc.

☐ Details

Methods - Data analysis

Which methods were used to analyse the data?

Please give details e.g. for in-depth interviews, how were the data handled? Details of statistical analysis can be given next.

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

- Scoring: maths pre- and post-tests were scored by calculating the number of questions correctly answered, which is consistent with the scoring scheme in several previous studies (Good et al. 2008; Johns et al. 2008; Spencer and Castano 2007; Steele and Aronson 1995). Highest possible score on each test was 15, which 3 participants achieved on the pre-test and 1 participant achieved on the post-test.
 - Scoring scheme for the working memory task was adapted from the AWMA scoring scheme. An answer on the working memory test was scored as correct if it was recalled in backwards order such that the first digit that participants heard was written last, the last digit first, and the intermediate digits were also in backwards order.
 - Survey responses: The free response question about what the research was studying was coded in binary as, “reported the stereotype relevant parts of the experiment”, or “did not report the stereotype relevant parts of the experiment”.

Which statistical methods, if any, were used in the analysis?

☐ Details

Preliminary analyses: - means and standard deviations - ANCOVA

Identity-based stereotype threat effects: -ANCOVAs with post-test scores as the dependent measure and pretest scores as the covariate.

Gender-based stereotype threat: - ANCOVA with gender as a between-subjects variable and maths pre-test score as a covariate.

Race-based stereotype threat: - ANCOVA with race as between-subject variable and maths pre-test score as a covariate - ANCOVA with working memory pre-test scores as the covariate

Income-based stereotype threat: - overall ANCOVA with income as a between-subject variable and maths pre-test controlled for - overall ANCOVA with income as a between-subject variable, working memory pre-test score as the covariate, and working memory post-test as the outcome - LSD post hoc comparisons - One-way between-subject ANOVA of self-reported income-based stereotype threat experienced during the experiment as a function of one's income was also run

Multiple minority stereotype threat: - ANCOVA on maths test performance as a function of the number of stigmatized aspects of identity (NSAI) that participants possessed when maths pre-test scores were covaried - post hoc analysis

Ancillary analysis: Effort: - mean scores - one-way between-subject ANOVA - Post hoc LSD comparisons

What rationale do the authors give for the methods of analysis for the study?

e.g. for their methods of sampling, data collection, or analysis.

☐ Details

For evaluation studies that use prospective allocation, please specify the basis on which data analysis was carried out.

‘Intention to intervene’ means that data were analysed on the basis of the original number of participants as recruited into the different groups. ‘Intervention received’ means data were analysed on the basis of the number of participants actually receiving the intervention.

- ☐ Not applicable (not an evaluation study with prospective allocation)
- ☐ ‘Intention to intervene’
- ☐ ‘Intervention received’
- ☐ Not stated/unclear (please specify)

Do the authors describe any ways they have addressed the reliability of data analysis?

e.g. using more than one researcher to analyse data, looking for negative cases.

- ☐ Details

Do the authors describe any ways they have addressed the validity of data analysis?

e.g. internal or external consistency; checking results with participants.

- ☐ Details

Do the authors describe strategies used in the analysis to control for bias from confounding variables?

- ☐ Details

Please describe any other important features of the analysis.

- ☐ Details

Please comment on any other analytic or statistical issues if relevant.

- ☐ Details

Results and Conclusions

How are the results of the study presented?

e.g. as quotations/figures within text, in tables, appendices.

- ☐ Details

- figures
- in text

What are the results of the study as reported by authors?

Please give details and refer to page numbers in the report(s) of the study where necessary (e.g. for key tables).

□ Details

Preliminary analyses: - significant difference in maths SAT scores as a function of number of stigmatized aspects of identity - The zero-SAI group and the one-SAI group scored significantly higher than the two-SAI group and the three-SAI group - Significant difference in SAT score as a function of gender, race, and income. - The stigmatized group had a lower mean score - Females scored lower than males and RM participants scored lower than White participants - High-income participants scored significantly higher than middle or low-income participants - Differences are not controlled for or covaried in future analyses because the use of SAT scores in a stereotype threat ANCOVA would directly violate the number of assumptions of the theory of ANCOVA

Identity-based stereotype threat effects: Gender-based stereotype threat: - ANCOVA with gender as a between-subject variable and maths pre-test score as a covariate revealed no stereotype threat effect on maths test performance - Significant effect of stereotype threat on working memory performance as revealed by an ANCOVA with working memory pre-test score as the covariate - Females and males had higher performance on the post-test than the pre-test but males improved more than females - Independent-samples t-test of self-reported endorsement showed no significant difference between males and female - This suggests that females were not consciously experiencing stereotype threat effects (because males and females did not differ significantly).

Race-based stereotype threat: - ANCOVA revealed significant stereotype threat effect on maths performance - White students slightly improved from pre-test to post-test whereas racial/ethnic minority students' performance was slightly decreased - ANCOVA showed a significant stereotype threat effect on working memory performance - All students had higher performance on the post-test than the pre-test, but White students improved more than RM students - Differences in self-report of stereotype threat experienced on the basis of race are not reported because White participants were not asked this question.

Income-based stereotype threat: - ANCOVA showed that there was a significant effect of income on maths test performance - LSD post hoc test revealed significant difference between low-income participants and high-income participants but there was no difference between low-income participants and middle-income participants or middle-income participants and high-income participants - The low-income participants did more poorly at post-test than pre-test, whereas high-income participants improved - Pattern is consistent with a stereotype threat effect among low-income participants relative to high-income participants but not relative to middle-income participants - Pattern also suggests that there maybe be a linear relation between income level and amount of stereotype threat experienced. - ANCOVA showed a significant effect of income on working memory. - LSD post hoc comparisons showed that the mean for the low-income group was significantly lower than the score for the middle-income group and the high-income group and that the middle- and high-income groups were not significantly different from one another - Low-income students

performed slightly worse at post-test compared to pre-test, whereas middle-income and high-income participants both improved quite a bit from pre-test to post-test - Pattern is consistent with a stereotype threat effect among low-income participants relative to both middle- and high-income participants. - One-way ANOVA of self-reported income-based stereotype threat experienced during the experiment as a function of one's income was also run - Significant effect of income on self-report of stereotype threat experienced for the three-income levels - Post hoc comparisons indicate that the mean score for the low-income condition was significantly higher than the mean score for the high-income condition, indicating greater feelings of threat - Middle-income mean score was not significantly different from either the high- or low-income group - These results suggest that low-income participants reported experiencing stereotype threat effects and that the extent to which participants reported experiencing income-based stereotype threat may be related to their income.

Multiple minority stereotype threat: - ANCOVA showed significant effect on maths test performance - Post hoc analysis showed that the effect was driven by the three-SAI group showing a significantly different change in performance than the zero-SAI group, the one-SAI group, and the two-SAI group, whereas the zero, one, and two groups did not differ from another. - Three-SAI group performed more poorly on the post-test than the pre-test, while the zero, one, and two SAI groups all improved slightly on the post-test - Significant effect on working memory. - post hoc analysis showed that the effect was driven by the three-SAI group showing a significantly different change in performance. - three-SAI group performed more poorly on the post-test than on the pre-test, while the others improved on the post-test

Ancillary analysis:

Effort: - ANOVA revealed significant difference between the reported effort of high-, middle- and low-income participants - Post hoc LSD comparisons show that low-income participants reported exerting significantly more effort - One-way ANOVA revealed significant difference in effort among the four different NSAI groups - Post hoc comparisons showed that groups with more than one stigmatized aspect of identity between them were significantly different from one another in the effort they reported putting forward - Three-SAI group put forth more effort than the zero or one SAI group but did not report exerting more effort than the two-SAI group. - Two stigmatized aspects of identity group put forth more effort than the zero stigmatized aspects of identity group but did not significantly more than the one or three-SAI groups.

Was the precision of the estimate of the intervention or treatment effect reported?

- CONSIDER:
 - Were confidence intervals (CIs) reported?
 - ☐ Yes
 - ☒ No
 - ☐ Can't tell

Are there any obvious shortcomings in the reporting of the data?

- ☐ Yes (please specify)
- ☐ No

Do the authors report on all variables they aimed to study as specified in their aims/research questions?

This excludes variables just used to describe the sample.

- ☐ Yes (please specify)
- ☐ No

Do the authors state where the full original data are stored?

- ☐ Yes (please specify)
- ☐ No

What do the author(s) conclude about the findings of the study?

Please give details and refer to page numbers in the report of the study where necessary.

- ☐ Details

Surprisingly, there was no evidence of gender-based stereotype threat effects on math performance. The female participants did not do any worse on the math tasks after hearing a prime intended to induce gender-based stereotype threat. There were gender-based stereotype threat effects on working memory such that women improved less on the working memory post-test than did men.

Interestingly, gender-based stereotype threat effects were observed on working memory. Specifically, both male and female participants performed better on the working memory post-test than the pre-test; however, male participants improved more than the female participants.

Race-based stereotype threat effects on math performance and working memory were found. That is, hearing a prime intended to produce race-based stereotype threat effects did indeed make RM participants perform more poorly than White participants. The existence of race-based stereotype threat effects on math performance was consistent with expectations, as it has been observed in previous research.

Race-based stereotype threat effects on math performance and working memory were found. That is, hearing a prime intended to produce race-based stereotype threat effects did indeed make RM participants perform more poorly than White participants. The existence of race-based stereotype threat effects on math performance was consistent with expectations, as it has been observed in previous research.

Consistent with expectations and previous research, there was evidence of income-based stereotype threat effects on math performance (Harrison et al. 2006). On the math task, low-income participants performed significantly worse on the post-test (with pre-test scores as a covariate) than did the high-income participants. In fact, low-income participants did worse on the math post-test than they had on the pre-test, while high-income participants improved and middle-income participants stayed the same.

Low-income participants also experienced stereotype threat effects on working memory. They performed worse on the working memory post-test than on the pre-test, whereas high- and middle-income participants all improved on the working memory post-test, likely due

to a practice effect. Importantly, this is the first research to document the existence of income-based stereotype threat effects on working memory.

The math performance of participants with three stigmatized aspects of identity decreased after hearing the stereotype-relevant primes, whereas math performance of all other participants improved slightly on the post-test. The same pattern was found on the working memory measures. That is, participants with three-SAIs exhibited a drop in working memory function after the prime, whereas participants in the zero, one and two-SAI groups did not experience stereotype threat effects and in fact showed a boost in performance on the working memory post-test.

Quality of the study - Reporting

Is the context of the study adequately described?

Consider your answer to questions: Why was this study done at this point in time, in those contexts and with those people or institutions? (Section B question 2) Was the study informed by or linked to an existing body of empirical and/or theoretical research? (Section B question 3) Which of the following groups were consulted in working out the aims to be addressed in the study? (Section B question 4) Do the authors report how the study was funded? (Section B question 5) When was the study carried out? (Section B question 6)

- ☒ Yes (please specify)
☐ No (please specify)

Are the aims of the study clearly reported?

Consider your answer to questions: What are the broad aims of the study? (Section B question 1) What are the study research questions and/or hypotheses? (Section C question 10)

- ☒ Yes (please specify)
☐ No (please specify)

Is there an adequate description of the sample used in the study and how the sample was identified and recruited?

Consider your answer to all questions in Methods on 'Sampling Strategy', 'Recruitment and Consent', and 'Actual Sample'.

- ☐ Yes (please specify)
☒ No (please specify)
 - No age

Is there an adequate description of the methods used in the study to collect data?

Consider your answer to the following questions in Section I: Which methods were used to collect the data? Details of data collection methods or tools Who collected the data?

Do the authors describe the setting where the data were collected? Are there other important features of the data collection procedures?

- ☒ Yes (please specify)
☐ No (please specify)

Is there an adequate description of the methods of data analysis?

Consider your answer to the following questions in Section J: Which methods were used to analyse the data? What statistical methods, if any, were used in the analysis? Who carried out the data analysis?

- ☐ Yes (please specify)
☒ No (please specify)
- context is sometimes missing or just indirectly mentioned

Is the study replicable from this report?

- ☐ Yes (please specify)
☒ No (please specify)

Do the authors avoid selective reporting bias?

(e.g. do they report on all variables they aimed to study as specified in their aims/research questions?)

- ☒ Yes (please specify)
☐ No (please specify)

Quality of the study - Methods and data

Are there ethical concerns about the way the study was done?

Consider consent, funding, privacy, etc.

- ☐ Yes, some concerns (please specify)
☐ No concerns

Were students and/or parents appropriately involved in the design or conduct of the study?

- ☐ Yes, a lot (please specify)
☒ Yes, a little (please specify)
☐ No (please specify)

Is there sufficient justification for why the study was done the way it was?

- ☒ Yes (please specify)
☐ No (please specify)

Was the choice of research design appropriate for addressing the research question(s) posed?

- ☒ Yes (please specify)
☐ No (please specify)

To what extent are the research design and methods employed able to rule out any other sources of error/bias which would lead to alternative explanations for the findings of the study?

e.g. (1) In an evaluation, was the process by which participants were allocated to or otherwise received the factor being evaluated concealed and not predictable in advance? If not, were sufficient substitute procedures employed with adequate rigour to rule out any alternative explanations of the findings which arise as a result? e.g. (2) Was the attrition rate low and if applicable similar between different groups?

- ☐ A lot (please specify)
☒ A little (please specify)
☐ Not at all (please specify)

How generalisable are the study results?

- ☐ Details

Weight of evidence - A: Taking account of all quality assessment issues, can the study findings be trusted in answering the study question(s)?

In some studies it is difficult to distinguish between the findings of the study and the conclusions. In those cases please code the trustworthiness of this combined results/conclusion. Please remember to complete the weight of evidence questions B-D which are in your review specific data extraction guidelines.

- ☐ High trustworthiness (please specify)
☒ Medium trustworthiness (please specify)
☐ Low trustworthiness (please specify)

Have sufficient attempts been made to justify the conclusions drawn from the findings so that the conclusions are trustworthy?

- ☐ Not applicable (results and conclusions inseparable)
☒ High trustworthiness
☐ Medium trustworthiness
☐ Low trustworthiness

Wells et al. (2014)

CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

Is the case definition adequate?

- a) yes, with independent validation
- b) yes, e.g., record linkage or based on self reports
- c) no description

Representativeness of the cases

- a) consecutive or obviously representative series of cases *
- b) potential for selection biases or not stated

Selection of Controls

- a) community controls *
- b) hospital controls
- c) no description

Definition of Controls

- a) no history of disease (endpoint) *
- b) no description of source

Comparability

Comparability of cases and controls on the basis of the design or analysis

- a) study controls for _____ (Select the most important factor.)
*
- b) study controls for any additional factor * (This criterion could be modified to indicate specific control for a second important factor.)

Exposure

Ascertainment of exposure

- a) secure record (e.g., surgical records) *
- b) structured interview where blind to case/control status *
- c) interview not blinded to case/control status
- d) written self report or medical record only
- e) no description

Same method of ascertainment for cases and controls

- a) yes *
- b) no

Non-Response rate

- a) same rate for both groups *
 - b) non respondents described
 - c) rate different and no designation
-

COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability.

Selection***Representativeness of the exposed cohort***

- a) truly representative of the average _____ (describe) in the community *
- b) somewhat representative of the average _____ in the community *
- c) selected group of users, e.g., nurses, volunteers
- d) no description of the derivation of the cohort

Selection of the non exposed cohort

- a) drawn from the same community as the exposed cohort *
- b) drawn from a different source
- c) no description of the derivation of the non exposed cohort

Ascertainment of exposure

- a) secure record (e.g., surgical records) *
- b) structured interview *
- c) written self report
- d) no description

Demonstration that outcome of interest was not present at start of study

- a) yes *
- b) no

Comparability***Comparability of cohorts on the basis of the design or analysis***

- a) study controls for _____ (select the most important factor) *
- b) study controls for any additional factor * (This criterion could be modified to indicate specific control for a second important factor.)

Outcome***Assessment of outcome***

- a) independent blind assessment *
- b) record linkage *
- c) self report
- d) no description

Was follow-up long enough for outcomes to occur

- a) yes (select an adequate follow up period for outcome of interest) *
- b) no

Adequacy of follow up of cohorts

- a) complete follow up - all subjects accounted for *
- b) subjects lost to follow up unlikely to introduce bias - small number lost - > _____ % (select an adequate %) follow up, or description provided of those lost) *
- c) follow up rate < _____ % (select an adequate %) and no description of those lost
- d) no statement

University of Glasgow (n.d.)

DOES THIS REVIEW ADDRESS A CLEAR QUESTION?***Did the review address a clearly focussed issue?***

- Was there enough information on:
 - The population studied
 - The intervention given
 - The outcomes considered

- ☐ Yes
- ☐ Can't tell
- ☐ No

Did the authors look for the appropriate sort of papers?

- The 'best sort of studies' would:
 - Address the review's question
 - Have an appropriate study design

- ☐ Yes
- ☐ Can't tell
- ☐ No

ARE THE RESULTS OF THIS REVIEW VALID?***Do you think the important, relevant studies were included?***

- Look for:
 - Which bibliographic databases were used

- Follow up from reference lists
- Personal contact with experts
- Search for unpublished as well as published studies
- Search for non-English language studies

- ☐ Yes
☐ Can't tell
☐ No

Did the review's authors do enough to assess the quality of the included studies?

- The authors need to consider the rigour of the studies they have identified. Lack of rigour may affect the studies results.

- ☐ Yes
☐ Can't tell
☐ No

If the results of the review have been combined, was it reasonable to do so?

- Consider whether:
 - The results were similar from study to study
 - The results of all the included studies are clearly displayed
 - The results of the different studies are similar
 - The reasons for any variations are discussed

- ☐ Yes
☐ Can't tell
☐ No

WHAT ARE THE RESULTS?

What is the overall result of the review?

- Consider:
 - If you are clear about the review's 'bottom line' results
 - What these are (numerically if appropriate)
 - How were the results expressed (NNT, odds ratio, etc)

How precise are the results?

- Are the results presented with confidence intervals?

- ☐ Yes
☐ Can't tell
☐ No

WILL THE RESULTS HELP LOCALLY?

Can the results be applied to the local population?

- Consider whether:
 - The patients covered by the review could be sufficiently different from your population to cause concern

- Your local setting is likely to differ much from that of the review

- ☐ Yes
- ☐ Can't tell
- ☐ No

Were all important outcomes considered?

- ☐ Yes
- ☐ Can't tell
- ☐ No

Are the benefits worth the harms and costs?

- Even if this is not addressed by the review, what do you think?
- ☐ Yes
- ☐ Can't tell
- ☐ No

References

- Critical Appraisal Skills Programme. (2018). CASP Systematic Review Checklist [Organization]. In *CASP - Critical Appraisal Skills Programme*. <https://casp-uk.net/casp-tools-checklists/>.
- EPPI-Centre. (2003). *Review guidelines for extracting data and quality assessing primary studies in educational research* (Guidelines Version 0.9.7). Social Science Research Unit.
- Tine, M., & Gotlieb, R. (2013). Gender-, race-, and income-based stereotype threat: The effects of multiple stigmatized aspects of identity on math performance and working memory function. *Social Psychology of Education, 16*(3), 353–376. <https://doi.org/10.1007/s11218-013-9224-8>
- University of Glasgow. (n.d.). *Critical appraisal checklist for a systematic review* [Checklist]. Department of General Practice, University of Glasgow.
- Wells, G., Shea, B., O'Connell, D., Robertson, J., Welch, V., Losos, M., & Tugwell, P. (2014). The newcastle-ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. *Ottawa Health Research Institute Web Site, 7*.