Rydell et al. (2014)

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
- executive functioning (Miyake et al., (2000)'s framework)
-

Do authors report how the study was funded?

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

- The author(s) received no financial support for the research, authorship, and/or publication of this article

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

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

- we argue that stereotype threat may affect several specific executive functions, which may account for different threat-related outcomes
- Specifically, we are interested in examining which executive function(s) can account for the stereotype threat's impact on women's outcomes within the stereotyped domain (maths performance) and outside of it (risk taking).

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

- women under stereotype threat

What is the relevant age group?

☐ Not applicate (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 applicate (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 cuntries was the study carried out?

☐ Explicitly stated (please specify)

☐ Not stated/unclear (please specify)

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

What are the study reserach 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)

- We predict that the executive function of updating accounts for threat's detrimental effect on women's maths performance, whereas inhibition accounts for women's increased risk preference when under threat effects.
- Because the skills involved in updating are likely impacted by stereotype threat and are also necessary to correctly solve difficult maths problems, we expect updating to account for women's maths performance under threat.
- We do not expect that updating will account for the effect of stereotype threat on risk-taking.

- We predict that reduced inhibition will account for increased risk taking under threat.
- We do not expect that inhibition will account for threat's impact on women's maths performance.
- We expected that stereotype threat would negatively affect inhibition and updating
- We expected that the negative effect of stereotype threat on updating would then account for reduced maths performance among women, whereas we expected the negative effect of threat on inhibition to account only for their increased risk taking.

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)

- inhibition
- shifting
- updating
- stereotype threat
- maths performance
- risk taking.

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)

☒ Prospecitive 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)

Experiment 1: - gender - stereotype threat vs control

Experiment 2: - stereotype threat vs control

Experiment 3: - stereotype theat vs control

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)

Experiment 1: - four

Experiment 2: - two

Experiment 3: - two

Was the assignment of participants to interventions randomised?

☐ Not applicable (not more than one group)

☐ Not applicate (no prospective allocation)

☒ Random

☐ Quasi-random

☐ Non-random

☐ Not stated/unclear (please specify)

Experiment 1: - random

Experiment 2: - random

Experiment 3: - random

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.

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)

- explicitly, women under stereotype threat

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)

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)

Were any incentives provided to recruit people into the study?

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

Experiment 1: - course credit

Experiment 2: - course credit

Experiment 3: - course credit or extra credit

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)

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)

Experiment 1: - Male (n = 93) and female (n = 75) undergraduates participated for course credit. They were randomly assigned to a control condition or a stereotype threat condition.

Experiment 2: - We recruited only female participants for Experiment 2 because men showed no effects of stereotype threat in Experiment 1. - Ninety (N = 90) female undergraduates, participating for course credit, were randomly assigned to a control condition or a stereotype threat condition.

Experiment 3: - Eighty-two (N = 82) female undergraduates participated for course credit or extra cfredit. They were randomly assigned to the control condition or the stereotype threat condition. Due to a computer error, data from one participant were incomplete, leaving the final sample of 81 women.

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)

Experiment 3: - n = 81, due to a computer error, data from one participant were incomplete.

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)

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)

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)

Experiment 1: - no information besides gender about the participants -> low

Experiment 2: - no information besides gender about the participants -> low

Experiment 3: - no information besides gender about the participants -> low

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

Experiment 1: - MA task -> b - stereotype threat manipulation -> b - Executive functions task -> b

Experiment 2: - Introduction and manipulation of stereotype threat similar to Experiment 1 - Executive functions task -> b - Maths performance

- Streotype threat manipulation similar to Experiment 1
- Executive functions task -> b
- Risk measure -> 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)

Experiment 1: - MA task: following Beilock et al., (2007)'s findings and tutorial - executive functions task: -> inhibition: Stroop (1935) task -> updating: letter-memory task (Morris & Jones, 1990) -> shifting: number-letter task - Stereotype threat manipulation: similar to Beilock et al., 2007

Experiment 2: - Maths performance: 10 min to complete 15 word problems taken from the GRE and. - Executive functions task: -> inhibition: Stroop (1935) task -> updating: track task (Miyake et al., 2000) -> shifting: colour-shape task
-> rest similar to Experiment 1

Experiment 3: - risk measure: from Inzlicht and Kang's (2010) study - manipulation check: three-item measure of threat-based concern (Marx, 2012) - Maths performance: Same GRE-based maths test used in Experiment 2 - Executive functions task: -> inhibition: antisaccade task -> shifting: colour-shape task -> updating: letter-memory task

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)

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

☐ Details

Experiment 1: - correlation analyses - *Executive functions task*: each of the executive functions tasks was submitted to a 2 (gender) x 2 (stereotype threat) ANOVA - *Maths performance*: Accuracy and maths reaction time were each submitted to a 2 (Gender) x 2 (ST) ANOVA - *Mediation Analysis*: Preacher and Hayes's (2008) bias-corrected bootstrapping procedure for models with multiple mediators.

Experiment 2: - correlation analyses - *Executive functions task*: separate independent samples t-tests - *Maths performance*: t-tests - *Mediation Analysis*: same as Experiment 1

Experiment 3: - correlation analyses - *Manipulation check*: t-test - *Executive functions task*: t-test - *Maths performance*: t-test - *Risk taking*: t-test - *Mediational Analysis*: -> Maths performance: same as Experiment 1 and 2 -> Risk taking: same as Experiment 1 and 2

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

Experiment 1: - In text - table - figure

Experiment 2: - in text - figure

Experiment 3: - in text - figure

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

Experiment 1:
- *Correlations between executive functions*: There was a significant positive correlation between the inhibition and updating measures. The correlation between the inhibition and shifting measures, and the updating and shifting measures, were not significant - *Executive functions task*: The results for the Stroop task showed a significant two-way interaction. Women showed poorer inhibition in the stereotype threat condition than in the control condition. Men's level of inhibition did not differ as a function of threat. The results for the letter-memory task also showed a two-way interaction. Women showed poorer updating under the stereotype threat condition than under the control condition. Men's level of updating did not differ in response to the threat manipulation. The results for the number-letter task showed no significant effects. Shifting was not affected by gender, stereotype threat, or their interaction.
- *Maths performance*: The results for accuracy showed the expected two-way interaction. Women showed poorer accuracy under the stereotype threat condition than under the control condition. Men's accuracy did not differ as a function of threat. Maths reaction times only showed a marginally significant main effect of gender. Women tended to take longer

to complete an MA problem than men. The lack of a two-way interaction of gender and stereotype threat on maths reaction time indicates that the maths performance results were not due to a speed-accuracy trade-off. - *Mediational Analysis*: This model simultaneously examined the indirect effects whereby the interaction of gender and stereotype threat predicted each of the three executive functions, which in turn predicted maths accuracy. In the model, the direct relation between the interaction of gender and stereotype threat still predicted maths accuracy when executive functions scores were taken into account, but it was reduced. Although the interaction of gender and stereotype threat was related to both the Stroop (inhibition) task and the letter-memory (updating) task, only the bias-corrected 95% confidence interval (CI) for the indirect effect of the letter-memory task on the relation between the interaction of gender and stereotype threat and maths performance did not include 0. The bias-corrected 95% CIs for the indirect effect of the Stroop task and the number-letter indicates that only updating, and not inhibition or shifting, mediated the relation between the interaction of stereotype threat and gender and maths performance.

Experiment 2: - *Correlations between executive functions*: The inhibition measure showed a marginally significant correlation with the updating measure, but was not correlated with the shifting measure. The updating measure was not correlated with the shifting measure. - *Executive functions task*: There was a significant effect of stereotype threat on the inhibition task. Women under threat showed poorer inhibition than women in the control condition. The results for the updating task also showed an effect of stereotype threat. Women showed poorer updating in the stereotype threat condition than in the control condition. The stereotype threat manipulation did not affect performance on the shifting task. - *Maths performance*: Women in the stereotype threat condition showed lower maths accuracy than did women in the control condition. Women under stereotype threat also showed a lower number correct than did those under the control condition. Women under the stereotype threat and control condition attempted roughly the same number of maths problems. - *Mediational Analysis*: For maths accuracy, the direct relation of stereotype threat still predicted maths accuracy when performance on the executive functions task was taken into account, but it was substantially reduced. Although stereotype threat was related to both the inhibition (Stroop) task and the updating (keep track) task, only the bias-corrected 95% CI for the indirect effect of the updating task on the relation between stereotype threat and maths accuracy did not include 0. The bias-corrected 95% CIs for the indirect effect of the inhibition and shifting tasks included 0. The model for number corrected showed that stereotype threat predicted the number correct when performance on the executive function task was taken into account. The bias-corrected 95% CI for the indirect effect of the updating task on the relation between stereotype threat and number correct barely included 0, and the bias-corrected 95% CIs for the indirect effect of the inhibition and shifting tasks also included 0. These results indicate that only updating, and not inhibition or shifting, mediated the effect of stereotype threat on women's maths accuracy. This same pattern was seen for number correct, but was weaker.

Experiment 3: - *Manipulation check*: The manipulation of stereotype threat was effective. Women in the stereotype threat condition showed greater threat-based concern than women in the control condition - *Correlations between executive functions*: The inhibition measure was unrelated to the updating measure, and the shifting measure. The updating

measure was not related to the shifting measure. - *Executive functions task*: There was a significant effect of stereotype therat on the inhibition task. Women under threat showed poorer inhibition than women under the control condition. The results for the updating task also showed an effect of stereotype threat. Women showed poorer updating in the stereotype threat condition than in the control condition. The stereotype threat manipulation did not affect performance on the shifting task. - *Maths performance*: Women in the stereotype threat condition showed poorer maths accuracy than those in the control condition. Women under threat also showed a lower number correct than did those under the control condition. Women in the stereotype threat and control conditions attempted an equivalent number of maths problems. - *Risk taking*: Women in the stereotype threat condition rated the riskier lottery as relatively more preferable than those in the control condition. Women under threat also provided their preference more quickly than women not experiencing threat. Risk taking and maths performance were not correlated. - *Mediation analyses*: -> Maths performance: For maths accuracy, the direct relation of stereotype threat still predicted the maths accuracy when performance on the executive functions tasks was taken into account, but it was reduced. Although stereotype threat was related to both the inhibition (antisaccade) task and the updating (letter-memory) task, only the bias-corrected 95% CI for the indirect effect of the updating task on the relation between stereotype threat and maths accuracy did not include 0. The bias-corrected 95% CIs for the indirect effect of inhibition and shifting tasks included 0. The model for number correct showed that stereotype threat predicted the number correct when performance on the executive function was taken into account. However, the bias-corrected 95% CI for the indirect effect of the updating task on the relation between stereotype threat and number correct barely included 0, and the bias-corrected 95% CIs for the indirect effect of inhibition and shifting included 0. These results indicate that only updating, and not inhibition or shifting, mediated the effect of stereotype threat on women's maths accuracy. As in Experiment 2, this same pattern was seen for number correct, but was weaker.
-> Risk taking: The direct relation of stereotype threat no longer predicted risk taking when performance on the executive functions tasks was taken into account. Most importantly, only the 95% bias-corrected CI for the inhibition measure did not include 0, whereas zero was included for both the updating measure and shifting measure. This analysis shows a significant indirect path whereby stereotype threat leads to changes in inhibition ability that account for the preference for a risky choice.

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

Experiment 1: Experiment 1 showed that stereotype threat reduced women's maths performance but had no effect on men's maths performance. The effect of stereotype threat on women's maths performance was partially accounted for by the executive function of updating; however, it was not explained by the executive functions of inhibition or shifting. These results provide a more detailed picture of the cognitive mechanisms that lead women experiencing threat to show poorer maths performance. This experiment begins to examine the executive functions through which threat operates to reduce women's maths performance, finding that difficulty updating seems to be integral to understanding why stereotype threat hurts women's maths performance.

Experiment 2: In Experiment 2, we again showed that stereotype threat's deleterious effect on women's maths performance is predicted by updating. Although stereotype threat redcued women's ability to inhibit prepotent responses, neither inhibition nor shifting accounted for the impact of threat on women's math accuracy. By utilizing different measures of updating and shifting and by using a different math task, we were able to provide evidence that the mediational role of updating in accounting for stereotype threat effects on women's performance was probably not due to aspects of the particular measures used to assess the executive functions or to strong similarities between the skills used in the maths task and the measure of updating.

Experiment 3: Experiment 3 replicated the results from the previous experiments by showing that stereotype threat reduced women's math accuracy by impairing updating. However, the effect of stereotype threat on increased risk taking was accounted for by threat's detrimental influence on inhibition. This work highlights that, even within the same experiment, stereotype threat's effect on distinct outcome measures can be uniquely accounted for by different executive functions. This experiment also demonstrates that the influence of stereotype threat on different outcomes can be due to the failures of distinct executive functions.

General Discussion: The present work replicated previous stereotype threat research and provided further evidence for the general theory that reduced executive functioning mediates the relationship between stereotype threat and reduced performance or increased risk taking. Although the executive functions of both inhibition and updating were negatively impacted by stereotype threat, only updating was found to account for women's threat-related decrements in maths performance.

These findings highlight how different aspects of executive functioning can account for different outcomes due to stereotype threat and speak to the benefits of utilizing research and theory on executive functioning to hone our predictions, as this work postulates how different executive functions can be delineated and why they are important for understanding cognitive functioning.

Our approach could predict which executive functions will account for outcome differences due to the particular aspects of the experimental tasks.

Furthermore, reduced inhibition in response to stereotype threat accounted for increased risk taking; this finding indicates that stereotype threat may lead to other types of risky behaviours such as risky sexual behaviours and drug and alcohol use.

This work also suggests a new framework within which interventions that aim to mitigate stereotype threat can be evaluated and applied. Specifically, testing how manipulations that have been previously shown to reduce threat (e.g., self-affirmation, "gender fair" information, learning about threat) have their influence on specific executive functions could aid in our theoretical understanding of how these manipulations work and provide practical suggestions as to under which conditions each way of mitigating threat may be most successful.

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)

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)

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.

Rydell, R. J., Van Loo, K. J., & Boucher, K. L. (2014). Stereotype threat and executive functions: Which functions mediate different threat-related outcomes? *Personality and Social Psychology Bulletin*, *40*(3), 377–390. https://doi.org/10.1177/0146167213513475

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*.