Krendl et al. (2008)

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

- Findings suggesting that working memory and performance anxiety both play a role in stereotype threat

Do authors report how the study was funded?

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

Support for this research was provided by National Science Foundation Grant SBE-035440, the Dartmouth Brain Imaging Center, and a Natural Science Foundation Graduate Fellowship to A.C.K.

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

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

It remains unclear how the mechanisms that are believed to underlie stereotype threat cause subsequent performance decrements. In order to effectively override the stereotype-threat phenomenon, it is vital to understand clearly the core processes that underlie it. In the current study, we used functional magnetic resonance imaging (fMRI) to identify the neural processes engaged when women perform difficult math tasks both in the present and in the absence of stereotype threat.

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

- Women performing math related tasks

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)

United States

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)

The researchers state that the current study used fMRI to identify neural processes engaged when women perform difficult math tasks both in the presence and in the absence of stereotype threat. But this is not a falsifiable hypothesis, and the authors do not state any specific research questions.

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)

- fMRI data
- math performance (math event data)
- stereotype threat

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)

☒ Explicityly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

They did only differ in their exposure to stereotype threat

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)

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)

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)

A male experimenter who was blind to condition recruited the participants, gave them generic instructions for the task, and put them in the scanner. During the experiment,

specific instructions that varied between conditions were presented by computer.

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)

Women under stereotype threat in math related tasks

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)

- Female undergraduate students were recruited for a study measuring "the neural mechanisms engaged in cognitive tasks that require both speed and accuracy."

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)

- Undergraduate women, who gave themselves a score of 4 or higher on a 7-point LIkert scale, regarding the sentence "It is important to me that I am good at math"

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)

Participants received partical course credit or $20 remuneration for participating.

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)

- N = 28; 14 control subjects

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)

- United States, further ethnic background or whether or not they were permanent residents, was not mentioned.

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)

- Implied by them being undergraduate students

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)

All participants were right-handed female undergraduates at Dartmouth College and were highly identified with math.

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

- Identification with math, response on a Likert scale -> a
- fMRI data -> b
- math performance (math event data) -> b

- stereotype threat -> 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)

- fMRI scanner: 3.0-T Philips Intera using an eight-channel phase arrayed coil.

- Visual stimuli were generated with an Apple G3 laptop computer running PsyScope software.
- They were projected onto a screen positioned at the head end of the bore by an Epson (Model ELP-7000) LCD projector
- A fiber-optic, light-sensitive key press interfaced with the PsyScope button box.
- fMRI data were analyzed using Statistical Parametric Mapping software.
- IAT
- math problems (e.g. "Is $19 * 6 - 6^2 = 78$?")
- ROIs tool in SPM99

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)

- Datmouth Brain Imaging Center

Are there other important features of data collection?

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

☐ Details

- The authors thank Bryan Denny for his assistance in data collection.

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

- ANOVA
- parameter estimates (β) and t-contrast images
- region-of-interest (ROI) analyses

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

The authors tank Emily Cross, David Kramer, and Brian Russ for their assistance with data analysis.

Results and Conclusions

How are the results of the study presented?

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

☐ Details

- Figures and tables in text, numbers 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

Behavioural Results *Overall Performance*: Performance (total number of math items answered correctly) was assessed in an ANOVA with time (Time 1 or Time 2) as within-subjects variable and condition (control or threat) as a between-subjects variable. The ANOVA revealed a significant interaction, and no main effect of time or condition. Subsequent t-tests indicated that the interaction emerged because the performance of control participants improved significantly over time, whereas the threatened group's performance decreased slightly over time.

Reaction Times: The mean log reaction times for each participant were then analyzed in a mixed-model ANOVA patterned after the analysis of overall performance. This ANOVA revealed a main effect of time, and no interaction.
Subsequent t-tests revealed that control participants' reaction times were significantly faster on the second math test than on the first. Participants in the threatened group, however, showed no significant difference in reaction time between the first and second tests. Reaction time was not associated with accuraccy.

fMRI Results From the imaging data, we identified neural activations that were greater during the second math test than the first. These analyses showed that the control participants recruited more left-lateralized activation in the inferior prefrontal cortex (Brodmann's area, BA, 47), let inferior parietal cortex (BA 40), and bilateral angular gyrus (BA 39) over time. By contrast, the threatened participants revealed greater activity in the ventral anterior cingulate cortex (vACC; BA 32/10) on the second test than on the first.
In order to determine if there were functional dissociations between these ROIs, we conducted a 2 (condition: control or threat) x 2 (time: Time 1 or Time 2) ANOVA for each ROI identified in the preceding paragraph. Significant interactions emerged for BA 47, left BA 39, and right BA 39, there was also a trend for BA 40. Post-hoc statistical tests revealed that the two-way interactions emerged because control participants recruited BA 48, BA 40, and BA 39 to a greater extent over time, whereas threatened participants did not.
A separate ANOVA conducted on vACC activation revealed a significant interaction, threatened participants, but not control participants, recruited heightened vACC activation over time.
These findings suggest a functional dissociation between vACC and the other ROIs. To formally test for this functional dissociation, we computed additional ANOVAs that included region as a third factor. Specifically, we conducted 2 (time) x 2 (condition) x 2 (region: vACC vs. each cognitive region) mixed-model ANOVAs with time and region as within-subjects variables and condition as a between-subjects variable. In these analyses, functional dissociations between regions could occur because (a) the regions behaved differently across conditions (resulting in a Region x Condition interaction), (b) the regions behaved differently across time (causing a Region x Time interaction), or (c) the regions demonstrated different Condition x Time interactions (resulting in a three-way interaction.)
We found a significant three-way interaction in each ANOVA - BA 47, left BA 40, left BA 39, right BA 39. The three-way interactions show that regions exhibiting a significant change in activation over time in one group of participants did not do so in the other group of participants.
Together, these interactions revealed a double dissociation between cognitive regions (left inferior frontal, left parietal, and bilateral angular gyrus regions) and an affective region (vACC). Whereas left inferior frontal, left parietal, and bilateral angular gyrus regions demonstrated increased activity for control participants but not threatened participants over time, vACC demonstrated the reverse pattern (i.e., increased activity for threatened but not control participants over time).

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

The results obtained are consistent with predictions from stereotype-threat literature. Women who were reminded of the stereotype threat that there are sex differences in math ability underperformed on subsequent math problems compared with women who received no such remained. These behavioural differences were accompanied by robust differences in neural activity between the two conditions.

Among women in the control condition, the brain regions that revealed heightened activation over time are consistent with the regions that previous imaging studies have identified as being associated with math computations, such as mathematical calculations, spatial representation of numbers, and mental rotation. These regions have been implicated in math learning.

The predominantly left lateralization of neural activation over time in our control participants is also consistent with previous research suggesting that right-handed participants rely on left prefrontal areas (i.e., inferior and medial frontal gyri) to perform mental math tasks and left lateralized parietal cortex when they have begun to develop an expertise on specific math tasks.

Together, these findings provide compelling evidence that participants in our control condition were developing strategies that enhanced their performance on the math task, as evidenced by their behavioural improvement over time.

Women who received the stereotype-threat reminder, however, failed to show increased recruitment of these mathematical brain regions and also performed worse over time. Unlike

the control condition, the threat condition was associated with robust activity in the vACC. Together, these findings shed light on the mental operations that underlie the pernicious effects of stereotype threat. Specifically, in the presence of stereotype threat, women experience heightened activation in the vACC, a region implicated in social and emotional processing, and subsequently fail to adopt effective mathematical learning strategies.

With respect to the predominant theories about the mechanisms that give rise to streotype threat effects, our finding soffer two important insights: First, stereotype threat does not necessarily increase demands on working memory (as evidenced by the fact that neural regions that support working memory were no more active after the threat induction than before). Second, a primary neural consequence of stereotype threat is to increase activation in the vACC, a region involved in social and emotional processing. We certainly are not suggesting that working memory is unaffected by stereotype threat. Rather, the heightened vACC activation we observed may occur at the expense of successfully recruiting other neural networks, such as regions involved in mathematical learning or working memory, as suggested by Beilock et al. (2007).

The primary effects of stereotype threat observed in this study may reflect several different psychological processes. For instance, stereotype threat may direct women's attention toward the negative social and emotional consequences of confirming negative stereotypes about their group, thereby increasing performance anxiety. Stereotype threat may also cause women to overmonitor their performance. Being reminded of negative stereotypes about their group may heighten women's awareness of their performance and limit the cognitive resources they devote to the task.

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)

The authors noted what they were looking into but never mentioned any hypothesis or research questions.

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)

- The age is missing as well as recruiting methods and the course they were taking. As Jończyk et al. (2022) noted, women majoring in fields that are mostly studied by men might be more resilient or have higher coping mechanisms, thus the course and age of the sample might actually end up being essential to know.

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)

- However, how the results are being displayed in text is suboptimal and could be better for the reader.

Is the study replicable from this report?

☐ Yes (please specify)

☒ No (please specify)

- Authors did not specify which math problems were used.

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)

- can't say since they did not specifically mention what they wanted to look into or report

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

- authors did not mention consent or the IRB approval

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)

- see Jończyk et al. (2022) for examples of potential bias.

How generalisable are the study results?

☐ Details

- not very, the sample isn't too specific, however the authors did not provide enough, and relevant information about the participants.

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

- The authors were very one-sided in their conclusions, and did not provide enough information to back up their claims.

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.

Jończyk, R., Dickson, D. S., Bel-Bahar, T. S., Kremer, G. E., Siddique, Z., & Van Hell, J. G. (2022). How stereotype threat affects the brain dynamics of creative thinking in female students. *Neuropsychologia*, *173*, 108306. https://doi.org/10.1016/j.neuropsychologia.2022.108306

Krendl, A. C., Richeson, J. A., Kelley, W. M., & Heatherton, T. F. (2008). The negative consequences of threat: A functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math. *Psychological Science*, *19*(2), 168–175. https://doi.org/10.1111/j.1467-9280.2008.02063.x

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*.