Jończyk et al. (2022)

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)

This study addresses whether, and if so how, exposure to a stereotype threat impacts the neural and behavioural correlates of creative thinking in female engineering students. In the present study, we used a population likely vulnerable to stereotype threat targeting creative thinking (undergraduate female engineering students) and asked them to perform two standard creative thinking tasks before and after exposure to a gender-related stereotype threat. If stereotype threat negatively impacts the ability to perform creatively - either in number of ideas produced (fluency) or the quality of ideas themselves (originality), an alpha decrease might be expected. Alternatively, if stereotype threat is actually stirring rather than discouraging, then stereotype threat delivery might lead to an alpha power increase along with improvement in creative thinking efforts (either in idea fluency or originality). Finally, buildinging on prior work repoting increased alpha power is associated with more original ideas, we also correlated alpha power to the idea of originality and fluency and expected a positive association based on past work.

Do authors report how the study was funded?

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

The project was supported by grants from the National Science foundation DUE 1561660 to Zahed Siddique and Janet van Hell, DUE IUSE 1726811 to Janet van Hell, DUE IUSE 1726884 Zahed Siddique, and DUE IUSE 1726358 to Gül E. Kremer and by a grant from the Foundation of Polish Science (START 41.2018) to Rafal Jonczyk.

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

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

Effects of stereotype threat on creative thinking and alpha power.

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

Female engineering undergraduate students, as they operate in a STEM field with a relatively large gender disparity.

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)

American students but the grant and author is polish

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)

If stereotype threat negatively impacts the ability to perform creatively - either in number of ideas produces (fluency) or in the quality of ideas themselves (originality), an alpha decrease might be expected.
Alternatively, if stereotype threat is acutally stirring rather than discouraging, then stereotype threat delivery might lead to an alpha power increase along with improvement in creative thinking efforts (either in idea fluency or originality).
Finally, building on prior work reporting increased alpha power is associated with more original ideas, we also correlated alpha power to the idea of originality and fluency and expected a positive association based on past work.

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)

- creativity -> fluency and originality
- alpha power

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)

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)

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)

Individulas 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)

Not stated, just said that they were female undergraduate students from a large Amrerican univeristy.

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)

Female undergraduate students majoring in engineering from a specific university.

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)

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)

All participants gave informed consent to participate in the experiment that was approved by the university's Institutional Review Board.
All participants were right-handed native speakers of English, had normal or corrected-to-normal vision, and reported no history of neurological impairment.

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)

22 female undergraduate students majoring in engineering (M_{age} = 19.1; SD = 0.89) - unknown if the last drop out (due to noisy data) changes the average and SD

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)

Twenty-seven female undergraduate students gave consent to the study, 23 ended up in the final sample of which one additional person was excluded due to noisy data.

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)

Stated that they are native speakers of English and that they attend a large American university, no further information given.

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)

M_{age} = 19.1; SD = 0.89, no specific age range stated.

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 native speakers of English, had normal or corrected-to-normal vision, and reported no history of neurological impairment.

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)

It is the exact sample the authors wanted to get -> female undergraduate students in engineering.

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

- demographic questionnaire (after consent).
- EEG data (alpha power)
- Creative thinking tasks (fluency and originality) -> specifically Alternate Uses task (AUT) and Utopian Situations task (UST)

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)

EEG -> alpha power
AUT and UST -> fluency and originality
demographic data

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

AUT and UST were tested in different research.

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)

Five independent and trained raters judged the originality of the generated ideas on a five-point Likert scale ranging from 1 (not original) to 5 (very original). Raters were instructed that a given answer was to be considered original when it was novel/unique and when it was principally possible. They were also asked to use the complete scale range as far as possible.

Where were the data collected?

e.g. school, home.

☐ Explicitly stated (please specify)

☒ Implicit (please specify)

☐ Unclear/not stated (please specify)

- lab due to the EEG

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)

R Core Team was used for all statistical analyses.

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

☐ Details

- indepentend and trained raters judged originality of the generated ideas on a five-point Likert scale.
- behavioural data analysis focused on fluency and originality of generated ideas in both the AUT and UST combined. The behavioural data were analyzed with two repeated measures ANOVAs with stereotype threat (pre-threat, post-threat) as a within-subject factor.
- EEG: IC cluster analysis, correlation analyses, TRP values in the lower and upper alpha band of the three alpha-related brain clusters were analyzed with an RM ANOVA, with STEREOTYPE THREAT (pre vs. post) as a within-subject factor. A Greenhouse-Geisser correction was applied where applicable and p-values obtained from post-hoc comparisons were adjusted using the Holm correction.

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

see above: UST and AUT rating was done by independent and trained raters.

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

In text, as 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

- idea originality did not differ between the post-threat and pre-threat conditions. The Pearson's product-moment correlation between idea originality and the pre-threat and post-threat conditions combined and openness to experience was positive but failed to reach statistical significance. Participants' score on the stereotype threat vulnerability scale did not correlate with the originality of ideas generated after the administration of stereotype threat.

- Idea fluency: The number of ideas did not differ between post-threat and pre-threat conditions. Participants' socre on the stereotype threat vulnerability scale did not correlate with the number of ideas generated after the administration of stereotype threat.
EEG

- Sensory-level analysis: In the lower alpha range, the ANOVA revealed a main effect of threat, with greater alpha Event-Related Synchronization (ERS) after the administration of stereotype threat. The main effect of hemisphere showed greater alpha ERS in the right compared to the left hemisphere. An area-by-hemisphere interaction showed greater alpha ERS in the right vs. left frontocentral, centrotemporal, centroparietal, and parietal areas. The threat-by-block interaction was not significant, ruling out the possibility that alpha power increased as a function of time on task. Finally, we compared alpha power in thne block directly preceding (block 2) and directly following (block 3) the stereotype threat. This one-way ANOVA was significant, and showed an increase in alpha power in block 3 rather than block 2. This result demonstrates that alpha power increased after the administration of stereotype threat.
  In the upper range, the RM ANOVA showed a main effect of threat, with greater upper alpha ERS after the administration of stereotype threat. The main effect of hemisphere, showed a greater alpha ERS in the right compared to left hemisphere. An area-by-hemisphere interaction, showed greater alpha ERS in the right vs. left centrotemporal, centroparietal, and parietal areas. The threat-by-block interaction was significant. Before the stereotype threat, alpha ERS was somewhat greater in block 1 than in block 2. By contrast, after the stereotype threat, greater alpha ERS was observed in block 4 rather than block 3. Finally, a direct comparison of upper alpha power in the block directly preceding (block 2) and directly following (block 3) the stereotype threat was significant and showed an increase in alpha power in block 3 rather than block 2. This result supports the finding from lower alpha range, demonstrating an increase in alpha power directly after the stereotype threat. Other comparisons did not differ from chance.

- Correlational analyses: Participants' ideational fluency did not correlate with alpha power in pre-threat and post-threat conditions either in the lower alpha range, or the upper alpha range. In the same vein, idea originality did not correlate with alpha

power in the pre-threat and post-threat conditions either in the lower alpha range, or in the upper alpha range.
For the left posterior alpha IC, the ANOVA showed a main effect of threat in the lower alpha range, with greater alpha ERS in the post-threat than in the pre-threat condition. Similarly, the effect of threat was significant in the upper alpha range with greater alpha ERS in the post-threat than in the pre-threat condition.
For the right posterior alpha IC, the ANOVA showed a main effect of threat in the lower alpha range, with greater alpha ERS in the administration of stereotype threat. In the upper alpha range, the main effect of threat only approached significance with slightly greater alpha ERS after the administration of stereotype threat.
Finally, for the central posterior alpha IC, the ANOVA showed a main effect of threat in the lower alpha range, with greater alpha ERS after the administartion of stereotype threat. Likewise, the effect of threat was significant in the upper alpha range with greater alpha ERS after the administration of stereotype threat.

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 primary finding in terms of brain activity was that alpha power increased post-threat relative to pre-threat, in both upper and lower alpha frequency bands and across

both sensor-level and independent component cluster-level analyses. The behavioural effects did not reach statistical significance but trended towards participants generating more ideas post-threat, with those ideas being less original than the pre-threat ideas had been.
These results appear consistent with an interpretation that the stereotype threat manipulation stirred and motivated participants to try harder, leading to increased internal attention to focuso n the task as reflected on alpha power, but that this effort did not lend itself to better behavioural performance.

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)

a comparison group that is unaffected by stereotype threat would've been nice. After all it could be possible that the results appear regardless of stereotype threat but instead due to the task at hand. Using a before and after is good but future research might also want to add a control group.

How generalisable are the study results?

☐ Details

The population, female engineering students, is probably familiar with the stereotype that women are not as good as men in STEM fields, however them getting this far might also mean that they are less affected by the stereotype - maybe they developed better coping skills or more resilience. Women with fewer coping skills might not make it to the major and drop out before hand, alternative, the participants might be less affected by the stereotype because they made it to the major - similar how the stereotype "women are bad at math" can be less impactful if the parents encourage their daughters to pursue math.

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)

highly trustworth for this specific group but not very generalizable to other groups.

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.

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

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*.