# Brown and Harkins (2016)

**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
- Working Memory Deficit Account:
- Mere Effort Account

***Do authors report how the study was funded?***

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

- This research was funded by the U.S. Army Research Institute for the Behavioral and Social Sciences (Contract W5J9CQ-12-C-0046;

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

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

- If stereotype threat participants withdraw effort on the intervening task when it is not described as related to gender-math stereotype, then it is possible that Mrazek et al. (2011) found poor SART performance (i.e., failures of controlled attention) for their stereotype threat participants because of differences in motivation rather than differences in working memory.

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

- individuals under gender-math 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)

- 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)

- In Mrazek et al.'s (2011) experiment, participants were given 2 s to provide their responses, providing ample time for correcting the prepotent response to press the response key. Given this, the mere effort account makes four predictions.

1. Participants in the GRE-stereotype threat/SART unrelated condition will perform more poorly than GRE-no stereotype threat/SART-unrelated participants, replicating Mrazek et al.'s (2011) findings.
2. In the SART-stereotype threat conditions, the difference in performance between GRE-stereotype and GRE-no stereotype threat participants should be eliminated, replicating Jamieson and Harkins (2011). That is, when performance on the SART is related to gender-maths stereotype, the participants in these two conditions should be equally and highly motivated to perform well on the SART, whether or not they anticipate taking the maths test under stereotype threat.
3. Participants in the SART-stereotype threat conditions should outperform in the SART-unrelated conditions whether they are assigned to the GRE-no stereotype threat condition
4. or to the GRE-stereotype threat condition.

**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)

- Stereotype threat manipulation
- Mind-wandering
- SART performance
- Response time
- Manipulation check
- GRE-problems

***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)

- Stereotype threat vs no stereotype threat condition

***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)

*Mrazek replication*: - GRE stereotype threat x SART no stereotype threat - GRE no stereotype threat x SART no stereotype threat

*No Mrazek replication*: - GRE stereotype threat x SART stereotype threat - GRE no stereotype threat x SART stereotype threat

***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)

- it is not mentioned how the participants were assigned to the GRE stereotype vs GRE no stereotype conditions.
- It is mentioned that the participants within the GRE stereotype threat and GRE no stereotype threat conditions were each split in half into SART stereotype threat vs SART no stereotype threat conditions -> Resulting in the 2 (GRE stereotype threat vs. GRE no stereotype threat) x 2 (SART stereotype threat vs. SART no stereotype threat) design.

***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.*

1. Greeted and told about math test
2. Stereotype threat manipulation vs no manipulation
3. 3 sample GRE problems (GRE stereotype threat vs GRE no stereotype threat)
4. SART manipulation (SART stereotype threat vs SART no stereotype threat)
5. SART task
6. two manipulation checks (one for SART and one for the math task, the math task were supposedly the GRE problems, which were never performed)
7. Experiment concluded after the two manipulation checks and no math task was given.

It should be noted that the description of the "procedure" is very ambiguous. It is said that in the Mrazek replication conditions, half of the participants were told that the SART is unrelated and the other half was told that the SART showed gender differences. This implies that they are still referring to the other half of the Mrazek replication conditions. But then we'd be left with the participants in "no replication conditions" and without any information about what they did or what they were told.
Thus the study only described the procedure for half of the participants.

However, it can also be implied (but then it is written in a very misleading way) that "half of the participants" refers to the total sample size. In this case the study design would be as described above, and a 2 x 2 design would be used. However, by the authors description it is not clear which of the two interpretations is correct. I just assume that it is the latter one since otherwise, the study does not make any sense.

**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)

***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)

- University students

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

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

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

***Planned sample size***

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

☐ Not applicable (please specify)

☒ Explicitly stated (please specify)

☐ Not stated/unclear (please specify)

- Power analysis with d = .70, power size of 80% resulted in N = 53 as a minimum total sample size.

**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)

- Course 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)

- Seventy-three female undergraduate students (N = 73)

***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)

- 100%

***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)

- The authors did not give relevant information about the participants, so it can't be determined.

***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

- response times -> b

- GRE stereotype threat -> b
- GRE no stereotype threat -> b
- SART stereotype threat -> b
- SART no stereotype threat -> b
- mind wandering -> b
- manipulation checks -> 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)

- GRE math problems
- SART task
- manipulation checks ("To what extent do you believe that gender differences exist on this task", 6 point Likert scale AND "who do you believe performs better on this task?", 11 point Likert scale AND "to what extent is performance on this task related to math ability?" 11 point Likert scale)

***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

- Manipulation Checks: 2 (anticipated threat: GRE-stereotype threat vs GRE-no stereotype threat) x 2 (SART description: SART-stereotype threat vs SART-unrelated) between subjects ANOVA
- SART performance: multivariate ANOVA (MANOVA) with all four measures of mind-wandering as the dependent variables (i.e., commission errors, omissions, anticipations, and RT coefficient of variation)
  - The four predictions from the mere effort account can be tested in a linear contrast
  - Mere effort: MANOVA using the contrast weights from before
    - univariate tests of these predictions (see above)
  - Working memory: second MANOVA using the contrast weights that tested the working memory deficit predictions

***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

- none

***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

- table
- 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

**Manipulation checks**: *SART manipulation check*: - SART stereotype threat participants believed that this task was related to math ability to a greater extent than SART unrelated participants. - SART-stereotype threat participants also reported that gender differences existed on the SART to a greater extent than SART-unrelated participants. - SART-stereotype threat participants reported that men performed better than women

on the SART to a greater extent than SART-unrelated participants. - The effect of anticipated threat and the interaction did not approach significance in any analysis. *Math test manipulation*: - Participants did not differ in their ratings of the extent to which performance on the anticipated math test reflected math ability - GRE-stereotype threat participants reported that the math test would produce gender differences to a greater extent than GRE-no stereotype threat participants. - GRE-stereotype threat participants also reported that men would outperform women on the upcoming math test to a greater extent than GRE-no stereotype threat participants. - The effect of SART description and the interaction did not approach significance in any analysis. - These results indicate that our manipulations were successful.

**SART performance**: *Statistical analysis of performance*: - The four predictions from the mere effort account can be tested in a linear contrast that predicts that the performance of participants in the SART-stereotype threat condition will not differ as a function of whether they anticipated taking the math test under stereotype threat or not and will exhibit less mind-wandering than participants in the GRE-no stereotype threat/SART-unrelated condition: - These three groups are also predicted to show less mind-wandering than participants in the GRE-stereotype threat/SART-unrelated condition -> These weights were chosen to create three levels of performance (i.e., low, moderate, high), in line with the predictions depicted in Figure 1 - The working memory account predicts that participants given the stereotype threat manipulation for the anticipated math test, the SART, or both will exhibit equivalent amount of mind-wandering, all more than the mind-wandering found in the GRE-no stereotype threat/SART-unrelated condition. - These weights map onto those depicted in Figure 1, such that the working memory account suggests that being assigned to a stereotype threat condition for either measure should lead to debilitated performance on the SART *Mere effort*: - There was a significant overall effect for the mere effort contrast - The univariate tests of these predictions also yielded significant effects supporting the mere effort account for commission errors, omissions, anticipations, RT coefficient of variation - The proportion of variance in the $SS_{between}$ that was explained by the contrast was calculated - For commission errors, the contrast testing the mere account predictions accounted for 98.7% of the variance between conditions. - For omissions, the contrast accounted for 75.8% of the variance - The mere effort predictions accounted for 71.8% of the variance for anticipations - 99.2% of the variance was captured by the contrast for RT coefficient variation - The residuals of this contrast were also tested to determine whether any systematic variance remained for commission errors, omissions, anticipations, and RT coefficients of variation - Means and SD for all four dependent variables are presented in Table 1 *Working memory*: - Experiencing stereotype threat for either the anticipated math test, the SART, or both, leads to equivalent performance, which is worse than that found in the GRE-no threat/SART-unrelated condition. - The multivariate test of this prediction was not significant - None of the univariate tests of this prediction were significant - The proportion of variance accounted for by the contrast was also calculated for these predictions - For commission errors, omissions, anticipations, and RT coefficient of variation, the contrast for the working memory deficit account predictions accounted for 0.4%, 11.5%, 14.5%, and 6.4% of the variance between groups, respectively - The residuals of this contrast were also tested to determine whether any systematic variance remained for commission

errors, omissions, anticipations, and RT coefficients of variation.

***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

- We replicated Mrazek et al.'s (2011) general pattern of findings on the SART with only minor differences
- However, we also found that across all four SART measures mere effort predicted substantially more variance than the working memory account
- Furthermore, nonsignificant tests of the mere effort contrast residuals suggest that these predictions accounted for all of the systematic variance in the means for three of the four SART measures, including commissions errors, whereas the tests of the working memory contrast residuals suggested that the data do not conform the predictions drawn from that perspective.

- This pattern of means suggests that when performance on the SART is introduced as unrelated to the manipulated stereotype, participants in the GRE-stereotype threat condition perform more poorly than their GRE-no stereotype threat counterparts.

- Describing the SARt as related to the same stereotype that was manipulated for the math test eliminated this difference between these two conditions
- Critically, in line with the suggestion that undergoing stereotype threat leads to increase motivation, participants in the SART-stereotype threat conditiosn also outperformed participants in each of the SART-unrelated conditions
- More specifically, the present research supports the motivational, mere effort account for stereotype threat performance effects over the cognitive, working memory deficit account. Knowing what causes stereotype threat performance effects is important because this knowledge is required for us to design interventions to counter these effects.

**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)

- NOPE, see above

***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)

- implicitly

***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)

- basically no description of the sample besides the gender

***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)

- can't tell

**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

- no mention of ethics or consent

***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)

- replication

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

☒ Yes (please specify)

☐ No (please specify)

- replication

***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

- can't tell, no description of the sample

***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

- this one is hard. The explanations for the results sound interesting but there might be a bias by the authors, since they also conducted the other studies that link the mere effort account and stereotype threat. Moreover, the mere effort account is based on the analysis of one of the authors, thus it is possible that they try to "defend" this stance.

- The explanations are also plausible and the papers that explain the performance issues under stereotype threat are also mostly by the same few researchers.
- It is unclear if the additions to the replication (SART stereotype threat) did not favour the mere effort account.

- Also i, personally, did not like the description of the study and its procedure at all, so i might be biased and more critical/negative than necessary.

**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

Brown, A. J., & Harkins, S. G. (2016). Threat does not make the mind wander: Reconsidering the effect of stereotype threat on mind-wandering. *Motivation Science*, *2*(2), 85–96. https://doi.org/10.1037/mot0000032

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.

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*.