Beilock et al. (2007)

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)

Study aim(s) and rationale

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

☒ Explicitly stated (please specify)

The study was informed by previous research on stereotype threat, working memory, and mathematical problem solving. The authors cite relevant work by Steele & Aronson (1995), Spencer et al. (1999), Schmader & Johns (2003), Ashcraft & Kirk (2001), and others.

Do authors report how the study was funded?

☒ Explicitly stated (please specify)

The research was supported by Institute of Education Sciences Grant R305H050004 to Sian L. Beilock and National Science Foundation Grant BCS-0601148 to Sian L. Beilock and Allen R. McConnell.

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

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

The study focuses on the cognitive mechanisms underlying stereotype threat effects in mathematical problem solving, particularly how stereotype threat impacts working memory resources.

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

The population focus is women, specifically female college students.

What is the relevant age group?

☒ 17 - 20

☒ 21 and over

The participants were undergraduate students, so likely included both 17-20 year olds and those 21 and over.

What is the sex of the population focus/foci?

☒ Female only

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

☒ Higher education institution

In Which country or cuntries was the study carried out?

☒ Explicitly stated (please specify)

The study was carried out in the United States, at two Midwestern universities.

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

The study examines how stereotype threat affects women's performance on mathematical problem solving tasks, specifically modular arithmetic problems. It investigates the impact of stereotype threat on working memory resources, especially phonological aspects of working memory. The study also explores ways to mitigate stereotype threat effects and examines potential spillover effects to other cognitive tasks.

What are the study reserach questions and/or hypotheses?

☒ Explicitly stated (please specify)

The main research questions are: 1. How does stereotype threat impact neural activation patterns and cognitive processes in mathematical problem solving? 2. What are the specific neural and cognitive mechanisms through which stereotype threat influences math performance?

The study tests several hypotheses across 5 experiments: 1. Stereotype threat will impair performance on math problems that rely heavily on working memory resources. 2. Stereotype threat will particularly impact problems that rely on phonological aspects of working memory. 3. Practicing math problems to automaticity will reduce stereotype threat effects. 4. Stereotype threat effects can spill over to subsequent tasks that rely on the same cognitive resources.

Methods - Design

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

☒ Explicitly stated (please specify)

Key variables measured include: - Math problem accuracy - Math problem reaction times - Working memory capacity - Phonological working memory load - Verbal and spatial working memory task performance - Self-reported thoughts/worries under stereotype threat

Study timing

☒ Cross-sectional

The study used a cross-sectional design, examining participants at a single time point across multiple experiments.

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

☒ Before and after

In most experiments, baseline performance was measured before the stereotype threat manipulation, and then performance was measured again after the manipulation.

Methods - Groups

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

☒ Prospecitive allocation into more than one group (e.g. allocation to different interventions, or allocation to intervention and control groups)

Participants were randomly assigned to stereotype threat or control conditions in several experiments.

How do the groups differ?

☒ Explicityly stated (please specify)

The groups differed based on exposure to a stereotype threat manipulation (informing participants about gender differences in math performance) versus a control condition.

Number of groups

☒ Two

Most experiments compared a stereotype threat group to a control group.

Was the assignment of participants to interventions randomised?

☒ Random

Participants were randomly assigned to conditions.

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

☒ Not stated/unclear (please specify)

The paper does not explicitly state whether allocation was concealed.

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

☒ Yes

Other than the stereotype threat manipulation, procedures appear to have been identical across groups.

Study design summary

This study used an experimental design across 5 experiments. Women were randomly assigned to stereotype threat or control conditions and completed mathematical problem solving tasks (modular arithmetic). Performance was compared between conditions. The study also examined working memory processes using dual-task paradigms and transfer effects to other cognitive tasks.

Methods - Sampling strategy

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

☒ Implicit (please specify)

The authors do not explicitly state they are aiming for a representative sample, but they use college student participants to examine stereotype threat effects in academic settings.

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

☒ Not stated/unclear (please specify)

The specific methods for identifying and recruiting participants are not clearly stated.

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

☒ Not stated/unclear (please specify)

The specific selection methods are not clearly stated, beyond noting that participants were undergraduate students who met certain criteria.

Planned sample size

☒ Not stated/unclear (please specify)

A planned sample size is not reported.

Methods - Recruitment and consent

Which methods are used to recruit people into the study?

☒ Not stated/unclear (please specify)

Specific recruitment methods are not described.

Were any incentives provided to recruit people into the study?

☒ Not stated/unclear (please specify)

The paper does not mention if incentives were provided.

Was consent sought?

☒ Participant consent sought

The paper mentions that participants provided informed consent.

Are there any other details relevant to recruitment and consent?

☒ No

Methods - Actual sample

What was the total number of participants in the study (the actual sample)?

☒ Explicitly stated (please specify)

The total sample across all 5 experiments was 143 women.

What is the proportion of those selected for the study who actually participated in the study?

☒ Not stated/unclear (please specify)

This information is not provided.

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

☒ Explicitly stated (please specify)

Participants were from the United States, specifically from two Midwestern universities.

What ages are covered by the actual sample?

☒ Not stated/unclear (please specify)

Exact ages are not reported, but participants were undergraduate students.

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

☒ Not stated/unclear (please specify)

Socioeconomic status is not reported.

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

☒ Not stated/unclear (please specify)

Ethnicity information is not provided.

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

☒ Not stated/unclear (please specify)

No information is provided about special educational needs.

Is there any other useful information about the study participants?

☒ Explicitly stated (please specify no/s.)

Participants were selected based on having at least moderate self-reported math skills and valuing math ability.

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

☒ Unclear (please specify)

Not enough information is provided to determine representativeness.

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

☒ Not applicable (e.g. study of policies, documents, etc)

This was not a longitudinal study.

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)

This was not a longitudinal study.

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)

This was not a longitudinal study.

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

(a) To define the sample: Self-reported math skills and importance of math ability were used to select participants.

(b) To measure aspects of the sample: Math problem accuracy and reaction times, working memory task performance, self-reported thoughts/worries.

Which methods were used to collect the data?

☒ Self-completion questionnaire
☒ Self-completion report or diary
☒ Psychological test

Details of data collection methods or tool(s).

☒ Explicitly stated (please specify)

The study used modular arithmetic problems as the main mathematical task. Working memory was assessed using dual-task paradigms and n-back tasks. Self-reported thoughts were collected using open-ended questionnaires.

Who collected the data?

☒ Not stated/unclear

The paper does not specify who collected the data.

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

☒ Details

The authors used established tasks (modular arithmetic, n-back) and counterbalanced problem types. They also removed poorly performed items to ensure reliability.

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

☒ Details

The authors cite previous research validating the use of modular arithmetic and n-back tasks. They also demonstrate the validity of their manipulations through pilot testing and manipulation checks.

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

☒ No (please specify)

There is no mention of concealment from those measuring outcomes.

Where were the data collected?

☒ Unclear/not stated (please specify)

The specific location of data collection is not stated, but it was likely in laboratory settings at the universities.

Are there other important features of data collection?

☒ Details

Performance feedback was provided in some experiments but not others. The order of conditions (baseline then experimental) was kept consistent across experiments.

Methods - Data analysis

Which methods were used to analyse the data?

☒ Explicitly stated (please specify)

The data were analyzed using analyses of variance (ANOVAs) and multiple regression analyses.

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

☒ Details

ANOVAs (including repeated measures), t-tests, multiple regression analyses.

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

☒ Details

The authors do not provide an explicit rationale for their choice of analyses, but the methods used are standard for experimental designs in psychology.

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

☒ Not stated/unclear (please specify)

The basis of data analysis is not explicitly stated, but appears to be 'intention to intervene' based on the reported analyses.

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

☒ Details

The authors report using 95% confidence intervals to assess significance for simple effects. They also removed poorly performed items to ensure reliability of the measures.

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

☒ Details

The authors conducted multiple experiments to replicate and extend their findings. They also used converging evidence from different measures (accuracy, reaction times, self-reports) to support their conclusions.

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

☒ Details

The authors counterbalanced problem types and used within-subjects designs in some experiments to control for individual differences. They also statistically controlled for reported thoughts/worries in some analyses.

Please describe any other important features of the analysis.

☒ Details

The authors log-transformed reaction times to reduce skew. They also used effect sizes (Cohen's d) to report the magnitude of effects.

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

☒ Details

The authors acknowledge the limitation of always having the baseline condition precede the experimental condition, but argue that this makes their effects more conservative.

Results and Conclusions

How are the results of the study presented?

☒ Details

Results are presented using a combination of in-text descriptions, tables of means and standard errors, and figures showing accuracy and reaction time data.

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

☒ Details

Key findings include: 1. Stereotype threat impaired performance on high-demand horizontal math problems but not low-demand or vertical problems. 2. Adding a phonological load impaired performance on horizontal but not vertical problems. 3. Extensively practicing math problems eliminated stereotype threat effects. 4. Stereotype threat effects spilled over to subsequent verbal (but not spatial) working memory tasks.

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

- CONSIDER:
  - Were confidence intervals (CIs) reported?

☒ Yes

The authors report 95% confidence intervals for key comparisons.

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

☒ No

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

☒ Yes (please specify)

The authors report on all variables mentioned in their research questions and hypotheses.

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

☒ No

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

☒ Details

The authors conclude that stereotype threat impairs math performance by co-opting verbal working memory resources needed for problem solving. They argue that stereotype threat induces verbal worries that compete for phonological resources. The authors also conclude that practicing problems to automaticity can alleviate stereotype threat effects, and that these effects can spill over to other tasks relying on verbal working memory.

Quality of the study - Reporting

Is the context of the study adequately described?

☒ Yes (please specify)

The authors provide a thorough background on stereotype threat research and working memory theories to contextualize their study.

Are the aims of the study clearly reported?

☒ Yes (please specify)

The aims and research questions are clearly stated in the introduction.

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

☒ No (please specify)

While basic information about the sample is provided, details about identification and recruitment are lacking.

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

☒ Yes (please specify)

The methods for each experiment are described in detail, including the tasks used and procedures followed.

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

☒ Yes (please specify)

The statistical analyses used are clearly described for each experiment.

Is the study replicable from this report?

☒ Yes (please specify)

The methods and procedures are described in sufficient detail to allow replication.

Do the authors avoid selective reporting bias?

☒ Yes (please specify)

The authors report on all variables and conditions mentioned in their methods.

Quality of the study - Methods and data

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

☒ No concerns

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

☒ No (please specify)

There is no mention of student or parent involvement in the study design.

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

☒ Yes (please specify)

The authors provide a clear rationale for their experimental design and choice of tasks based on previous research.

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

☒ Yes (please specify)

The experimental design with manipulation of stereotype threat and working memory demands was appropriate for addressing the research questions.

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?

☒ A lot (please specify)

The study uses multiple experiments with converging methods to rule out alternative explanations. They use control conditions, counterbalancing, and within-subjects designs to control for potential confounds. However, the fixed order of baseline and experimental conditions in some experiments is a limitation.

How generalisable are the study results?

☒ Details

The results may generalize to women in academic settings, particularly in math-related domains. However, generalizability is limited by the use of a specific math task (modular arithmetic) and a sample of college students from Midwestern universities.

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

☒ Medium trustworthiness (please specify)

The study uses rigorous experimental methods and provides converging evidence across multiple experiments. However, there are some limitations in sampling and the fixed order of conditions in some experiments. Overall, the findings can be trusted with moderate confidence.

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

☒ High trustworthiness

The authors provide thorough discussion of their findings, linking them to previous research and theory. They acknowledge limitations and provide multiple lines of evidence to support their conclusions.

Weight of evidence - B: Appropriateness of research design and analysis for addressing the question, or sub-questions, of this specific systematic review.

☒ High

The experimental design with manipulation of stereotype threat and working memory demands is highly appropriate for addressing the review questions about the impact of stereotype threat on neural activation patterns and cognitive processes in academic settings.

Weight of evidence - C: Relevance of particular focus of the study (including conceptual focus, context, sample and measures) for addressing the question, or sub-questions, of this specific systematic review.

☒ High

The study's focus on the cognitive mechanisms of stereotype threat in math problem solving is highly relevant to the review questions. The use of working memory measures and neural activation patterns (inferred from behavioral data) aligns well with the review's focus.

Weight of evidence - D: Overall weight of evidence

☒ High

Considering the appropriate design, relevant focus, and rigorous methods, this study provides strong evidence for addressing the review questions, despite some limitations in sampling and generalizability.

References

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*(2), 256–276. https://doi.org/10.1037/0096-3445.136.2.256