Lin et al. (2023)

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)

- Perspective taking
- Spatial tasks
- stereotype threat

Do authors report how the study was funded?

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

- This research was supported by the National Nature Science Foundation of China (31200778) and the Swedish Research Council.

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

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

- In this study, female university students were recruited as participants to address the effect stereotype threat has on females' spatial perspective taking directly.

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

- Female university students

What is the relevant age group?

☐ Not applicate (focus not learners)

☐ 0 - 4

☐ 5 - 10

☐ 11 - 16

☐ 17 - 20

☐ 21 and over

☐ Not stated/unclear

What is the sex of the population focus/foci?

☐ Not applicate (focus not learners)

☒ Female only

☐ Male only

☐ Mixed sex

☐ Not stated/unclear

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

☐ Community centre

☐ Correctional institution

☐ Government department

☒ Higher education institution

☐ Home

☐ Independent school

☐ Local education authority

☐ Nursery school

☐ Other early years setting

☐ Post-compulsory education institution

☐ Primary school

☐ Residential school

☐ Secondary school

☐ Special needs school

☐ Workplace

☐ Other educational setting

In Which country or cuntries was the study carried out?

☐ Explicitly stated (please specify)

☐ Not stated/unclear (please specify)

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

What are the study reserach questions and/or hypotheses?

Research questions or hypotheses operationalise the aims of the study. Please write in authors' description if there is one. Elaborate if necessary, but indicate which aspects are reviewers' interpretation.

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

- We hypothesized that gender stereotype threat impairs females' spatial perspective taking. That is to say, females who were primed by a negative stereotype threat would

perform worse on a spatial perspective-taking task than those who were not primed with a stereotype threat.

- We also aimed to explore the underlying mechanism of the effect of gender stereotype threat on females' spatial perspective-taking.
- Combining the impact of stereotype threat on inhibition and the important role of inhibition in spatial perspective taking, we predicted that reduced inhibition may account for the effect of stereotype threat on females' spatial perspective-taking ability.

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)

Experiment 1: - stereotype threat - spatial perspective taking - gender identification as a covariate variable

Experiment 2: - stereotype threat - spatial perspective taking - executive function tasks

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 control 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)

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)

- female students

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)

- 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)

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)

- ¥10

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 procedures performed in studies invovling human participants were in accordance with ethical standards of the univery research committee and with the 1964 Helsinki declaration

Methods - Actual sample

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

If more than one group is being compared please give numbers for each group.

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

Experiment 1: - A total of 76 undergraduate students (mean age = 18.36 years, SD = 1.17) were randomly recruited as participants in return for ¥10. - All participants were females (38 in the control condition and 38 in the stereotype threat condition) - There was no significant difference in age between the stereotype threat group (M = 18.32, SD = 1.25) and the control group (M = 18.39, SD = 1.10)

Experiment 2: - Eighty-four female undergraduate students participated in this study for ¥10. - Seven participants were excluded from the analyses due to equipment malfunction (n = 1), not completing the perspective-taking test as required (n = 2), or extreme outliers on executive function measures (n = 4). - The remaining 77 (mean age = 18.53, SD = 1.1.7) participants were included in the analyses - There was no significant difference in age between the conditions.

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

Please specify numbers and percentages if possible.

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

Experiment 2: 77, Seven participants were excluded from the analyses due to equipment malfunction (n = 1), not completing the perspective-taking test as required (n = 2), or extreme outliers on executive function measures (n = 4).

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)

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

If the study involves more than one group, please give drop-out rates for each group separately. If necessary, refer to a page number in the report (e.g. for a useful table).

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Not applicable (not following samples prospectively over time)
☐ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear

For studies that involve following samples prospectively over time, do the authors provide any information on whether and/or how those who dropped out of the study differ from those who remained in the study?

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Not applicable (not following samples prospectively over time)
☐ Not applicable (no drop outs)
☐ Yes (please specify)
☐ No

If the study involves following samples prospectively over time, do authors provide baseline values of key variables such as those being used as outcomes and relevant socio-demographic variables?

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Not applicable (not following samples prospectively over time)
☐ Yes (please specify)
☐ No

Methods - Data collection

Please describe the main types of data collected and specify if they were used (a) to define the sample; (b) to measure aspects of the sample as findings of the study?

☐ Details

Experiment 1: - Demographic information -> a - gender identification scale -> b - stereotype threat manipulation -> b - spatial perspective-taking task -> b - manipulation check -> b

Experiment 2: - demographic information -> a - gender identification scale -> b - stereotype threat manipulation -> b - executive function tasks (Stroop task, local-global task, keep track task) -> b - spatial perspective-taking task -> b - manipulation check -> b

Which methods were used to collect the data?

Please indicate all that apply and give further detail where possible.

☐ Curriculum-based assessment
☐ Focus group
☐ Group interview
☐ One to one interview (face to face or by phone)
☐ Observation
☐ Self-completion questionnaire
☐ Self-completion report or diary
☐ Exams
☐ Clinical test
☐ Practical test
☐ Psychological test
☐ Hypothetical scenario including vignettes
☐ School/college records (e.g. attendance records etc)
☐ Secondary data such as publicly available statistics
☐ Other documentation
☐ Not stated/unclear (please specify)

Details of data collection methods or tool(s).

Please provide details including names for all tools used to collect data and examples of any questions/items given. Also please state whether source is cited in the report.

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

Experiment 1: - *Gender identification*: "Importance to Identity" dimension in the Collective Self-Esteem Scale (Luhtanen & Crocker, 1992) - *Stereotype threat manipulation*: briefing report stating that perspective taking is a spatial ability in which males have an advantage over females and then portrayed the aim of the current study as that of measuring people's spatial ability - *Stereotype threat - neutral condition*: introduced basic information about poplar trees and portrayed the aim of the current study as that of measuring people's reading ability. - *Perspective-Taking Test*: Pencil-and-paper perspective-taking test, which was created by Hegarty and Waller (2004). - *Manipulation Check*: Participants were asked to indicate how much they agreed with three statements, which were adapted from items used in Marx's (2012) study

Experiment 2: - Materials for gender identification, spatial perspective-taking test and manipulation check were identical to those used in Experiment 1 - *Inhibition - executive function*: Stroop (1935) task - *Shifting 0 executive function*: Local-global task (Miyake et al., 2000) was modified to measure shifting. In this test, we used the "Navon figure" (Navon, 1977) - *Updating - executive function*: Keep track task (Yntema, 1963), was modified to measure the updating ability according to some recent studies. The experimental materials were changed to Chinese words with a high frequency of use.

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)

Experiment 1: - lab

Are there other important features of data collection?

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

☐ Details

Methods - Data analysis

Which methods were used to analyse the data?

Please give details e.g. for in-depth interviews, how were the data handled? Details of statistical analysis can be given next.

☐ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

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

☐ Details

Experiment 1: - *Manipulation Check*: independent t-test - *Gender Identification*: independent t-test - *Spatial Perspective-Taking Performance*: single factor (stereotype threat vs control) between-subjects ANOVA on the angular deviation in the spatial perspective-taking task by including gender identification as a covariate. Post-hoc power analysis

Experiment 2: - *Manipulation Check*: independent t-test - *Gender Identification*: independent t-test - *Executive Function Tasks*: single factor (stereotype threat vs control) between-subject ANCOVA with inhibition, updating, shifting, and spatial perspective-taking performance as dependent variables and gender identification as a covariance variable. Post-hoc power analysis - *Spatial Perspective-Taking Performance*: single factor (stereotype threat vs control) between-subjects ANOVA on the angular deviation in the spatial perspective-taking task by including gender identification as a covariate. Post-hoc power analysis - *Correlations between Variables*: Pearson product-moment correlation coefficients - *Mediation Analysis*: Preacher and Hayes' (2008) bootstrapping procedure for models with multiple mediators for examining mediation with small samples was employed.

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

e.g. for their methods of sampling, data collection, or analysis.

☐ Details

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

'Intention to intervene' means that data were analysed on the basis of the original number of participants as recruited into the different groups. 'Intervention received' means data were analysed on the basis of the number of participants actually receiving the intervention.

☐ Not applicable (not an evaluation study with prospective allocation)
☐ 'Intention to intervene'
☐ 'Intervention received'
☐ Not stated/unclear (please specify)

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

e.g. using more than one researcher to analyse data, looking for negative cases.

☐ Details

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

e.g. internal or external consistency; checking results with participants.

☐ Details

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

☐ Details

Please describe any other important features of the analysis.

☐ Details

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

☐ Details

Results and Conclusions

How are the results of the study presented?

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

☐ Details

Experiment 1: - in text

Experiment 2: - in text - figure - tables

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

Please give details and refer to page numbers in the report(s) of the study where necessary (e.g. for key tables).

☐ Details

Experiment 1: - *Gender Identification*: Independent t-test showed that there was no significant difference in gender identification between the stereotype threat and the control group. Thus activating the stereotype threat was not affected by participants' gender identification. - *Spatial Perspective-Taking Performance*: The degree of deviation from the correct direction (angular error) differed significantly between the stereotype threat condition and the control condition. The angular error was higher in the stereotype threat condition than in the control condition. A post-hoc power analysis using G*Power indicated that we were sufficiently powered to detect a medium-sized effect. This result suggested that females' performance on the perspective-taking task was undermined when stereotype threat was activated.

Experiment 2: - *Manipulation Check*: Independent t-test showed that the score of stereotype threat group was significantly higher than that of the control group. Therefore the manipulation of stereotype threat was effective in Experiment 2 - *Gender Identification*: There was no significant difference in gender identification between the stereotype threat group and the control group. - *Inhibition*: There was a significant effect of stereotype threat on the inhibition task. Females under threat showed lower inhibition than females in the control condition, indicating that stereotype threat reduced the inhibition function. Post-hoc power = .92 - *Updating*: The results for the updating task also showed a significant effect of stereotype threat. Females showed poorer updating performance in the stereotype threat condition than in the control condition, indicating that stereotype threat also negatively impacted the updating function. The post-hoc power analysis indicated that power was 0.66 - *Shifting*: There was no significant effect of stereotype threat on shifting. Thus, the shifting function was not affected by stereotype threat. - *Spatial Perspective-Taking Performance*: Females in the stereotype threat condition showed higher angular deviation than females in the control condition. This suggests that the females' spatial perspective-taking performance was affected by stereotype threat. The results replicated the results obtained from Experiment 1. Post-hoc power = .97 - *Correlations between Variables*: There was a significant positive correlation between the inhibition and spatial perspective-taking measures. The correlation between the shifting and the perspective-taking measures as well as the correlation between the updating and the perspective-taking measures were not significant. We also found that inhibition was correlated with updating but was not correlated with shifting. Updating was not correlated with shifting. - *Mediation Analysis*: The results showed that although stereotype threat was related to both the inhibition task and the updating task, only the mediation effect of the inhibition task on the relationship between stereotype threat and perspective taking was significant. The mediating effects of updating and shifting were not significant. We found that the direct effect of stereotype threat on the angular deviation of spatial perspective-taking test was statistically significant. Stereotype threat also had significant negative effects on both inhibition and updating, but no significant effect on shifting. Inhibition negatively predicted the angular deviation of the

spatial perspective-taking test. The effects of updating and shifting on the spatial perspective-taking test were not significant. Stereotype threat not only directly affected females' spatial perspective-taking ability but also damaged it through the inhibition function.

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

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

☐ Yes
☒ No
☐ Can't tell

Experiment 2: in mediation analysis it was reported

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

https://osf.io/kuymd/?view_only=764a2a0472a14107b7b6a6b48edc3321

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

Please give details and refer to page numbers in the report of the study where necessary.

☐ Details

Experiment 1: Findings obtained from Experiment 1 were fully consistent with our prediction that females would perform worse on the spatial perspective-taking task in the stereotype threat condition. This revealed that stereotype threat had a negative effect on female perspective-taking ability.

Experiment 2: The results of Experiment 2 verified the negative impact of stereotype threat on females' spatial perspective-taking, replicating the results obtained from Experiment 1. The results also showed that stereotype threat negatively impacted the inhibition and updating components of executive function. The results of the multiple mediation analysis indicated that inhibition mediated the relationship between stereotype threat and females' spatial perspective-taking. That is, the negative impact of stereotype

threat on spatial perspective-taking ability was partically caused by reduced inhibition. In addition, stereotype threat decreased the updating ability, but neither updating nor shifting mediated the effect of stereotype threat. These findings further detailed the cognitive mechanism of the effect of stereotype threat on spatial perspective-taking performance.

General Discussion: The current study explored the effect of stereotype threat on females' spatial perspective taking and identified a possible cognitive mechanism.
Our results showed that stereotype threat impaired females' spatial perspective taking. These results were consistent with previous research about the relationship between stereotype threat and other spatial abilities. Our research on spatial perspective-taking is meaningful and enriches the research structure in the field of stereotype threat and spatial ability.

Our study found that stereotype threat impairs executive functions. Specifically, the inhibition and updating components of executive function were lower in the stereotype threat condition than in the control condition. In earlier studies, executive function has been treated as a single variable without consideration of its components or has been one-sidedly understood as inhibitory control. Executive function has both a unified and a diverse structure. Recent research has focused on the relationship between stereotype threat and different components of executive function. Consistent with our findings, previous studies have found that stereotype threat could cause a decline in inhibition function and the updating function proposed that stereotype threat led to poor performance via depleted executive resources. The probable reason is that a negative stereotype threat leads to a reduction in self-regulation or an increase in negative emotion, resulting in a lowering of inhibition and updating. With respect to shifting, to our knowledge no research has shown that stereotype threat reduces the shifting function. Individuals might engage in a large amount of shifting to try to get out of the threat. Thus, shifting should not be impaired by stereotype threat.

It is worth noting that some researchers have been unable to replicate the effect of stereotype threat on the inhibitory control.

Although inhibition and updating were negatively affected by stereotype threat, only the decline in the inhibition function was related to the decline in females' spatial perspective-taking ability. Rydell et al. (2014) suggested that the impact of stereotype threat on different outcomes may be mediated by different executive functions. Our experimental results support this perspective in that the impact of stereotype threat on spatial perspective-taking was not mediated by the entire set of executive functions.

Conclusion: Overall, our results show that stereotype threat has a negative impact on females' performance in spatial perspective-taking ability. And the present study provides further evidence for the role of executive function. Specifically, stereotype threat reduced females' inhibition and updating function, rather than their shifting function. The inhibition function plays a mediating role in the relationship between stereotype threat and females' spatial perspective taking.

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)

- could also include the origin and ethnicity, since stereotype threat might differ in different cultures.

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

Consider your answer to the following questions in Section I: Which methods were used to collect the data? Details of data collection methods or tools Who collected the data? Do the authors describe the setting where the data were collected? Are there other important features of the data collection procedures?

☒ Yes (please specify)
☐ No (please specify)

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

Consider your answer to the following questions in Section J: Which methods were used to analyse the data? What statistical methods, if any, were used in the analysis? Who carried out the data analysis?

☒ Yes (please specify)
☐ No (please specify)

Is the study replicable from this report?

☒ Yes (please specify)
☐ No (please specify)

Do the authors avoid selective reporting bias?

(e.g. do they report on all variables they aimed to study as specified in their aims/research questions?)

☒ Yes (please specify)
☐ No (please specify)

Quality of the study - Methods and data

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

Consider consent, funding, privacy, etc.

☐ Yes, some concerns (please specify)
☒ No concerns

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

☒ Yes, a lot (please specify)
☐ Yes, a little (please specify)
☐ No (please specify)

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

☒ Yes (please specify)
☐ No (please specify)

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

☒ Yes (please specify)
☐ No (please specify)

To what extent are the research design and methods employed able to rule out any other sources of error/bias which would lead to alternative explanations for the findings of the study?

e.g. (1) In an evaluation, was the process by which participants were allocated to or otherwise received the factor being evaluated concealed and not predictable in advance? If not, were sufficient substitute procedures employed with adequate rigour to rule out any alternative explanations of the findings which arise as a result? e.g. (2) Was the attrition rate low and if applicable similar between different groups?

☐ A lot (please specify)
☒ A little (please specify)
☐ Not at all (please specify)

How generalisable are the study results?

☐ Details

- for females under stereotype threat (maybe just for Chinese)

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

In some studies it is difficult to distinguish between the findings of the study and the conclusions. In those cases please code the trustworthiness of this combined results/conclusion. ***Please remember to complete the weight of evidence questions B-D which are in your review specific data extraction guidelines.***

☐ High trustworthiness (please specify)
☒ Medium trustworthiness (please specify)
☐ Low trustworthiness (please specify)

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

☐ Not applicable (results and conclusions inseparable)
☒ High trustworthiness
☐ Medium trustworthiness
☐ Low trustworthiness

Wells et al. (2014)

CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

Is the case definition adequate?

- a) yes, with independent validation
- b) yes, e.g., record linkage or based on self reports
- c) no description

Representativeness of the cases

- a) consecutive or obviously representative series of cases *
- b) potential for selection biases or not stated

Selection of Controls

- a) community controls *
- b) hospital controls
- c) no description

Definition of Controls

- a) no history of disease (endpoint) *
- b) no description of source

Comparability

Comparability of cases and controls on the basis of the design or analysis

- a) study controls for _____ (Select the most important factor.) *
- b) study controls for any additional factor * (This criterion could be modified to indicate specific control for a second important factor.)

Exposure

Ascertainment of exposure

- a) secure record (e.g., surgical records) *
- b) structured interview where blind to case/control status *
- c) interview not blinded to case/control status
- d) written self report or medical record only
- e) no description

Same method of ascertainment for cases and controls

- a) yes *
- b) no

Non-Response rate

- a) same rate for both groups *
- b) non respondents described
- c) rate different and no designation

COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability.

Selection

Representativeness of the exposed cohort

- a) truly representative of the average _____ (describe) in the community *
- b) somewhat representative of the average _____ in the community *
- c) selected group of users, e.g., nurses, volunteers
- d) no description of the derivation of the cohort

Selection of the non exposed cohort

- a) drawn from the same community as the exposed cohort *
- b) drawn from a different source
- c) no description of the derivation of the non exposed cohort

Ascertainment of exposure

- a) secure record (e.g., surgical records) *
- b) structured interview *
- c) written self report
- d) no description

Demonstration that outcome of interest was not present at start of study

- a) yes *
- b) no

Comparability

Comparability of cohorts on the basis of the design or analysis

- a) study controls for _____ (select the most important factor) *
- b) study controls for any additional factor * (This criterion could be modified to indicate specific control for a second important factor.)

Outcome

Assessment of outcome

- a) independent blind assessment *
- b) record linkage *
- c) self report
- d) no description

Was follow-up long enough for outcomes to occur

- a) yes (select an adequate follow up period for outcome of interest) *
- b) no

Adequacy of follow up of cohorts

- a) complete follow up - all subjects accounted for *
- b) subjects lost to follow up unlikely to introduce bias - small number lost - > _____ % (select an adequate %) follow up, or description provided of those lost) *
- c) follow up rate < _____% (select an adequate %) and no description of those lost
- d) no statement

University of Glasgow (n.d.)

DOES THIS REVIEW ADDRESS A CLEAR QUESTION?

Did the review address a clearly focussed issue?

- Was there enough information on:
  - The population studied
  - The intervention given
  - The outcomes considered

☐ Yes
☐ Can't tell
☐ No

Did the authors look for the appropriate sort of papers?

- The 'best sort of studies' would:
  - Address the review's question
  - Have an appropriate study design

☐ Yes
☐ Can't tell
☐ No

ARE THE RESULTS OF THIS REVIEW VALID?

Do you think the important, relevant studies were included?

- Look for:
  - Which bibliographic databases were used
  - Follow up from reference lists
  - Personal contact with experts
  - Search for unpublished as well as published studies
  - Search for non-English language studies

☐ Yes
☐ Can't tell
☐ No

Did the review's authors do enough to assess the quality of the included studies?

- The authors need to consider the rigour of the studies they have identified. Lack of rigour may affect the studies results.

☐ Yes

☐ Can't tell
☐ No

If the results of the review have been combined, was it reasonable to do so?

- Consider whether:
  - The results were similar from study to study
  - The results of all the included studies are clearly displayed
  - The results of the different studies are similar
  - The reasons for any variations are discussed

☐ Yes
☐ Can't tell
☐ No

WHAT ARE THE RESULTS?

What is the overall result of the review?

- Consider:
  - If you are clear about the review's 'bottom line' results
  - What these are (numerically if appropriate)
  - How were the results expressed (NNT, odds ratio, etc)

How precise are the results?

- Are the results presented with confidence intervals?

☐ Yes
☐ Can't tell
☐ No

WILL THE RESULTS HELP LOCALLY?

Can the results be applied to the local population?

- Consider whether:
  - The patients covered by the review could be sufficiently different from your population to cause concern
  - Your local setting is likely to differ much from that of the review

☐ Yes
☐ Can't tell
☐ No

Were all important outcomes considered?

☐ Yes
☐ Can't tell
☐ No

Are the benefits worth the harms and costs?

- Even if this is not addressed by the review, what do you think?

☐ Yes
☐ Can't tell
☐ No

References

Critical Appraisal Skills Programme. (2018). CASP Systematic Review Checklist [Organization]. In *CASP - Critical Appraisal Skills Programme*. https://casp-uk.net/casp-tools-checklists/.

EPPI-Centre. (2003). *Review guidelines for extracting data and quality assessing primary studies in educational research* (Guidelines Version 0.9.7). Social Science Research Unit.

Lin, Y., Zhang, B., Jin, D., Zhang, H., & Dang, J. (2023). The effect of stereotype threat on females' spatial perspective taking and the mediating role of executive functions. *Current Psychology*, *42*(6), 4979–4990. https://doi.org/10.1007/s12144-021-01849-7

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*.