Johns et al. (2008)

EPPI-Centre (2003) & Critical Appraisal Skills Programme (2018)

If the study has a broad focus and this data extraction focuses on just one component of the study, please specify this here

☐ Not applicable (whole study is focus of data extraction)

☒ Specific focus of this data extraction (please specify)

- Experiment 1, 3 and 4
- 2 Excluded since no stereotype threat manipulation

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
- Stereotype threat and performance
- Interplay of cognitive and emotional processes
-

Do authors report how the study was funded?

☐ Explicitly stated (please specify)

☒ Implicit (please specify)

☐ Not stated/unclear (please specify)

This research was supported in part by National Institute of Mental Health Grant 1R01MH071749 awarded to Toni Schmader

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

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

- The current research was designed to fill this gap: Advances have been made in delineating the processes by which this threat operates, the field still lacks clear empirical tests of how different affective and cognitive processes work together to impair task performance.
- We examined whether targets of negative stereotypes underperform in threatening intellectual environments because they dedicate limited cognitive resources to controlling their emotional reactions to such situations.
- Specifically, we propose that being the target of negative stereotype can hurt performance because regulating one's anxious response to the situation hijacks the executive resources necessary for performing well in such domains.
- The purpose of this research was to examine whether targets of stereotype threat allocate cognitive resources to the process of regulating anxiety using suppression.

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

- not mentioned

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)

Experiment 1:
- We expected that stereotype-threatened participants would automatically attend to anxiety-related stimuli, providing evidence of increased anxiety when the measure was implicit. - In a second condition, we described the reaction time measure as an instrument designed to assess anxiety and provided information about the logic of the measure. In this condition, we expected that this information would allow targets to engage in *expressive suppression* by actively redirecting their attention away from threat-related stimuli. - We predicted that we would detect evidence of active suppression when stereotype threat was present and when the women were aware that their anxiety levels were being assessed. Because awareness that anxiety is being measured is not necessary for the experience of stereotype threat, we expected that stereotype threat would deplete executive resources regardless of whether or not participants were aware that anxiety was being assessed.

Experiment 3:
- If reappraising anxiety under threat buffers performance, it would provide more direct evidence that antecedent-focused coping reduces stereotype threat by eliminating anxiety suppression tendencies. Together, these procedures allow for a more precise and stringent test of the hypothesis that stereotype threat depletes executive resources via emotional self-regulation - The primary predication was that women under stereotype threat who were not given any reappraisal information would show lower working memory and maths performance compared to women in the other two conditions.

Experiment 4:
- We predicted that only Latino students in the no-reappraisal (i.e., threat-only) condition would show evidence expressive suppression and depleted working memory resources because Caucasian participants should not experience threat-based anxiety that they would feel motivated to regulate, their attention allocation and working memory resources should not be affected by the description of the study and the reappraisal manipulation. - In contrast, if stereotype threat increases concern with suppressing one's anxiety, then this motivation should be reduced if not eliminated among Latinos instructed to reappraise anxiety as benign. - As a result, Latinos in the reappraisal condition should show patterns of attention allocation and working memory equivalent that of Caucasians in this 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)

Experiment 1:
- Anxiety - Stereotype threat - Suppression of anxious responses - working memory task - self report of anxiety

Experiment 3:
- stereotype threat - reappraisal of anxiety - control - working memory - GRE maths/problem-solving problems - self-reported anxiety

Experiment 4:
- demographic information - stereotype threat / anxiety manipulation - rest similar to those used in the anxiety measure condition of study 1 - intelligence - analytical reasoning questions - dot probe - working memory task

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)

Experiment 1:
- 2 (stereotype threat) x 2 (anxiety measure description) between-subjects.

Experiment 3:
- 3 (stereotype threat only vs. stereotype threat and anxiety reappraisal vs. control)

Experiment 4:
- 2 (ethnicity) x 2 (effect of anxiety)

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.

Experiment 1: 1. upon entering the lab, all participants were seated at adjacent computer workstations
- in the stereotype threat manipulation, the female participant was seated at the middle workstation so that she was flanked by two male confederates - Male experimenter explained the purpose of this study was to administer a test of mathematical aptitude in order to collect normative data on men and women - in the no-threat condition, 3 female participants were told by a female experimenter that the purpose of the study was to administer a problem-solving exercise in order to collect normative data on college students 2. All participants were told that they would compelte the maths test/problem-solving exercise in two parts, separated by two filler tasks, and that they would receive performance feedback at the end of the session - the filler tasks were actually tasks of interest: the dot probe task and working memory measure.
3. Completion of initial set of word problems (to induce threat and bolster the cover story) 4. Dot probe task 5. experimenter announced that there would not be sufficient time to complete the second problem set 6. probing for suspicion, debriefed, and thanked for their participation.

Experiment 3:
1. Male experimenter conducted each session, which included a female participant and a confederate 2. Explanation 3. review of five maths problems 4. filler task: working memory task from study 1 5. 20 min to complete the problem-solving/maths test 6. questionnaire, measuring anxiety (same as study 1) and two manipulation checks

Experiment 4:
1. 2 caucasian female experimenters conducted the sessions two- to four-person groups that always included at least 1 Caucasian and 1 Latino participant 2. explanation 3. intelligence test (in two parts) 4. stereotype threat manipulation (through short set of analytical reasoning questions) 5. dot probe task (measure of anxiety) 6. Working memory task (described as a filler task) 7. questionnaire, measures of self-reported anxiety and manipulation checks

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)

Experiment 1, 3:
- women under maths stereotype threat

Experiment 4:
- Latino and Caucasian 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)

- 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)

Experiment 1:
- participants were recruited if they reported scoring at least 500 on the quantitative section of the SAT (or equivalent converted ACT score) and reported awareness of the relevant stereotype (i.e., responded 3 or below to this question "Regardless of what you personally believe, do you think there is a stereotype that men and women differ in their maths ability?")

Experiment 3: - Participants were recruited if they reported having knowledge of the stereotype about women's maths ability and being at or above the scale of midpoint on a measure of maths identification.

Experiment 4:
- Participants were recurited on the basis their self-reported ethnicity.

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)

- University

Were any incentives provided to recruit people into the study?

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

Experiment 1, 4:
- course credit or $10

Experiment 3:
- 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)

Experiment 1:
- 85 Caucasian female psychology students attending a large American university who participated in exchange for course credit or $10. - Four participants were excluded. - All analyses were conducted on a final sample of 81 women.

Experiment 3:
- 61 Caucasian women attending a medium-sized American university who completed the study for course credit.

Experiment 4:
- 34 Latino (22 women, 12 men) and Caucasian (28 women, 19 men) undergraduates who participated for course credit or $10.
- 3 Latino and 3 Caucasian participants excluded - Final sample of 31 Latinos and 44 Caucasians.

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 1:
- 81 out of 84

Experiment 3:
- 58 out of 61

Experiment 4: - 31 Latinos and 44 Caucasians out of 34 Latinos and 47 Caucasians

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)

Experiment 1, 3:
- Caucasian

Experiment 4: - Latino and Caucasian

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)

Experiment 1:
- Four women were excluded from analyses due computer malfunction (n = 1), a failure to follow task instructions (n = 2), or for knowing a member of the study personnel (n = 1). All analyses were conducted on a final sample of 81 women.

Experiment 3:
- Data from 3 participants were excluded due to procedural errors (n = 2) or prior completion of the working memory measure (n = 1), leaving a final sample of 58 women.

Experiment 4:
- A computer error resulted in data loss from 1 Latino participant, and 5 additional participants (2 Latino, 3 Caucasian) failed to follow the instructions on the computer tasks.

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: - reported SAT score -> a - reported awareness of the stereotype -> a - dot probe task -> b - working memory task -> b - self-reported anxiety -> b - demographic information -> a

Experiment 3:
- question about knowledge of the stereotype -> a - maths identification scale -> a - working memory task -> b - GRE maths/problem-solving problems -> b - self-reported anxiety -> b - demographic information -> a

Experiment 4:
- ethnicity -> a - intelligence test -> b - dot probe task -> b - working memory task -> b - self-reported anxiety -> b - demographic information -> a - manipulation checks -> b - analytical reasoning questions -> 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: - Self-reported anxiety: participants' ratings of how agitated, anxious, nervous, uneasy, and worried they felt using a 7-point scale - working memory: dual-processing measure of working memory, called the reading span task (Schamder & Johns, 2003). - dot probe task: anxiety and suppression of anxious responses using the dot probe task (Mathews & MacLeod, 1986) - Stereotype threat manipulation: using procedures Inzlicht and Ben-Zeev (2000) developed and validated

Experiment 3:
- problem solving/maths task: 30 multiple choice word problems taken from the quantitative section of the GRE (Schmader & Johns, 2003) - post test questionnaire: same as study 1 + two manipulation checks of anxiety reappraisal: "According to the researcher, how does anxiety affect performance on the types of problems you just completed?" and "How do you think anxiety affects performance on the types of problems you just completed?" Participants responded on a 7-point scale - working memory task: same as study 1

Experiment 4:
- analytical reasoning questions to elicit stereotype threat: shown to produce stereotype threat for Latinos (Gonzales, Blanton, & Williams, 2002; Schmader & Johns, 2003) - dot probe task: same as study 1 - working memory task: same as study 1 - self-reported anxiety: same as study 1 - manipulation checks: same as study 1 - intelligence test

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

Experiment 1: - 2 (stereotype threat) x 2 (anxiety measure description) between-subjects ANOVA on the absolute span score - analysis of attention - means and SE - correlation analyses

Experiment 3:
Manipulation checks: - averaged the two manipulation checks before analysing them in a one-way ANOVA
Working memory: - we analysed the number of words recalled on the working memory task using a set of orthogonal contrasts
Math test performance: - same analysis was conducted on the number of items answered correctly on the maths test
Self-reported anxiety: - see study 1
Mediation: - additional analyses to assess the degree to which the influence of the manipulation on maths test performance was associated with parallel changes in working memory - contrasted the performance - bootstrapping - product coefficient

Experiment 4:
Manipulation check: - 2 (reappraisal) x 2 (ethnicity) ANOVA on the anxiety information manipulation check
Expressive Suppression and Working Memory: - Tested a priori predictions with orthogonal contrasts, first tested the primary prediction that Latinos in the threat-only control condition would show evidence of expressive suppression and the lowest working memory compared to the Latinos in the anxiety reappraisal condition and Caucasians on both conditions - Second contrast compared performance of Caucasians with the threat-only control condition to Caucasians and Latinos in the anxiety reappraisal condition - Third contrast tested the simple main effect of ethnicity within the anxiety reappraisal condition
Self-reported anxiety: - see study 1

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: - figure - in text

Experiment 3, 4:
- in text - table

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: *Working memory*: - our prediction about the effect of emotion regulation on executive functioning translates into a main effect of stereotype threat on the number of words recalled on the working memory task - ANOVA yielded only the predicted main effect of stereotype threat - Women in the stereotype threat condition recalled fewer words compared to women in the problem-solving control condition.
Dot Probe Task: - Analysis of attention allocation yielded only the predicted interaction between stereotype threat and anxiety measure description - In the stereotype threat condition, when the dot probe task was described as a measure of perceptual focus, women directed more attention towards anxiety-related words, compared when the task was described as a measure of anxiety. - women in the problem-solving condition did not show differential attention towards anxiety related words when the task was described as a measure of perceptual focus compared to when the task was described as a measure of anxiety. - When the dot probe task was described in neutral terms, allocation of attention to threat-related stimuli was negatively correlated with working memory when under stereotype threat but uncorrelated with working memory in the control condition - When the dot probe task was described as a measure of anxiety, working memory was positive correlated with attention allocation to anxiety words for women under stereotype threat but uncorrelated for women in the control condition
Self-reported anxiety: - did not yield any significant effects - The overall average of self-reports was significantly lower than the scale midpoint - In the stereotype threat condition, self-reported anxiety did not correlated with responses on the dot probe task when it was described as a measure of perceptual focus or anxiety.

Experiment 3:
Manipulation checks: - Participants in the stereotype threat plus reappraisal condition reported that anxiety had a more benign effect on performance compared to participants in the threat-only condition who were not given any information about the effect of anxiety on performance and participants in the non threat control condition - In addition to confirming the effectiveness of the manipulation, these results indicate that participants not given information about the influence of anxiety assumed that it would have a greater negative effect on performance, which conceivably could provide the motivation to suppress anxiety in order to perform well.
Working Memory: - Participants in the threat-only condition recalled significantly fewer words compared to participants in the control and reappraisal conditions, who did not differ from one another.
Maths test performance: - As predicted, participants in the stereotype threat condition answered significantly fewer questions correctly compared to participants in the control and reappraisal condition, who did not differ from one another.
Self-reported anxiety: - Anxiety ratings did not differ by condition, and the average anxiety rating was significantly lower than the scale midpoint - If the reappraisal manipulation

improved performance by reducing the experience of threat-induced anxiety then we might have expected significantly lower levels of self-reported anxiety in that condition. - The fact that the reappraisal manipulation did not reduce self-reported anxiety suggests that interpreting anxiety as inert did not reduce stereotype threat by reducing the experience of anxiety.
Mediation: - The manipulation contrast had a significant effect on the proposed mediator, working memory - The number of words recalled on the working memory task was marginally related to maths test performance after controlling for the effect of the manipulation contrast. - When controlling for the relationship between working memory and maths test performance, the manipulation contrast was no longer a significant predictor of maths test performance - The indirect effec was .23, with a 95% CI ranging from .02 to .78 - Because the CI does not include zero, the indirect effect is significant at .05

Experiment 4:
Manipulation check: - ANOVA revealed only the expected main effect of the manipulation - Regardless of ethnicity, participants in the reappraisal condition had a more benign view of anxiety compared to those in the threat-only condition - There was a general tendency for participants in the threat-only condition to assume that feeling anxious would harm performance.
dot probe task: - Results of the first contrast confirmed that Latinos in the threat-only control condition directed less attention to anxiety-related words compared to participants in the other three conditions - Residuals from this contrast did not vary across conditions, suggesting that this contrast pattern describes the majority of between-groups variance in attention allocation - Replicating the pattern observed with women in Study 1, this result indicates that Latinos were attempting to suppress anxiety when no reappraisal information was provided.
- Directing participants to reappraise anxiety as unrelated to performance appeared to prevent Latinos from suppressing anxiety - results of the second contrast confirmed that the attention allocation of Caucasians in the threat-only control condition did not differ from the attention allocation of both Lations and Caucasians in the anxiety reappraisal condition - The third contrast indicated that there was no difference between the attention allocation of Latinos and Caucasians in the anxiety reappraisal condition.
Working memory: - When the same analysis was conducted on the absolute span score, results from the first contrast confirmed that Latinos in the threat-only control condition recalled significantly fewer words compared to participants in the other three conditions - The residuals for this condition did not vary across conditions - The second contrast reveals a tendency for Caucasians in the threat-only control condition to recall more words compared to Latinos and Caucasians in the anxiety reappraisal condition - This effect appears to be driven by unexpectedly low word recall scores by Caucasians in the anxiety reappraisal condition - The third contrast indicated that word recall for Latinos and Caucasians in the anxiety reappraisal condition did not differ significantly from one another - As in studies 2 and 3, a manipulation designed to prevent stigmatized targets from engaging in response-focused coping eased the cognitive burden that is typically observed under stereotype threat. - Furthermore, reappraisal had no benefits for a nontargeted group.
Self-reported anxiety: - There were no significant effects on self-reported anxiety.

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 3:
- yes

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

Experiment 1: *Dot probe task*: - Women under stereotype threat were experiencing increased anxiety but attempted to suppress their expression of anxiety when they were aware that they dot probe task measured this state - It was not the case that women showed a general tendency to shift attention away from anxiety-related stimuli when they thought that their anxiety was being assessed, rather stereotype threat appears to instigate efforts to suppress the expression anxiety. - Women under stereotype threat were in fact feeling anxious and that those feeling corresponded to reduced working memory efficiency - Women under stereotype threat who thought the task measured anxiety directed their attention away from threat-related stimuli to avoid appearing anxious, and the more they tried to regulate their anxiety, the fewer executive resources they had available.
Self-reported anxiety: - The lack of differences in self-reported anxiety, along with the lack of correlation between the implicit and explicit measures, is consistent with previous studies showing a dissociation between direct and indirect measures of the psychological experience of stereotype threat - Such dissociation would be expected if participants experiencing stereotype threat suppress the expression of anxiety as part of an attempt to regulate their

experience of this negative emotion.

‘ The results of this study offer preliminary support for the hypothesis that regulating anxiety through suppression contributes to the effect of stereotype threat on executive resource depletion. Taken together, the overall pattern of responses in this study indicates that stereotype threat was anxiety provoking but that it also motivated a desire to avoid displaying those anxious feelings. The fact that responses on the reaction time measure corresponded to reductions in working memory efficiency in the stereotype threat condition suggests further that emotion regulation efforts deplete the executive resources of targets.

Experiment 3:

Within the context of the process manipulation, these analyses provide further evidence that anxiety regulation contributes to the effect of stereotype threat on performance depleting executive resources (MacKinnon, Lockwood, & Williams, 2004).

Experiment 4:

dot probe task: These results demonstrate that expressive suppression is specific to targets of stereotype threat. Directing participants to reappraise anxiety as irrelevant for performance appears to reduce attempts among stigmatized targets to suppress anxious responses while under stereotype threat. Importantly, Caucasians showed no sign of such suppression tendencies even though they were given the same information that the purpose of the study was to examine the relationship between anxiety and performance on an intelligence test. Given that Caucasians showed no evidence of emotion-focused suppression in the threat-only condition, it is not surprising that the anxiety reappraisal instruction did not moderate their responses on this measure.

Taken together, the results from this study highlight that reappraisal is an effective means of buffering executive resources from the deleterious effects of stereotype threat because it reduces efforts to engage in response-focused suppression. The current study also suggests that the results of the prior studies would not apply to nonthreatened individuals under the conditions we have created: Only threatened Latino participants showed evidence of trying to avoid expressing anxiety and consequently benefited from the reappraisal manipulation. Finally, by replicating the results of the prior studies with a different stigmatized group, we also provide evidence suggesting that emotion regulation is a general reaction to stereotype threat and not specific to women.

General Discussion: Our results provide evidence that stereotype threat operates via the interplay between cognitive and affective processes. The results of these studies converge to show that during stereotype threat, targets experience anxiety that they try to regulate through suppression. This very act of emotion regulation—characterized here as response-focused coping—seizes on the same cognitive resources needed for the central task (e.g., taking a math test) and can result in suboptimal performance.

The results from the four experiments develop a chain of causality (Spencer et al., 2005) to provide converging evidence that stereotype threat depletes executive resources via emotion-focused regulation.

Study 1 established that targets of stereotype threat not only experience anxiety but that situations of threat cue attempts to monitor and suppress their expression of these feelings. Study 3 further demonstrated the benefits of reappraisal by instructing participants to

reappraise not the situation, but the anxiety they felt as a result of the situation. The results of this study provide further evidence that executive resources are preserved when targets are given an opportunity to deal with their emotions in a way that does not involve response-focused regulation.
Finally, Study 4 generalized these findings to a different stigmatized group and provided evidence that suppression tendencies are specific to stereotyped group members.
It is noteworthy that our results converge across studies that vary considerably in methodology. We manipulated stereotype threat in four different ways and reappraisal in two different ways, examined two different stigmatized groups, both measured and manipulated suppression tendencies, and used two different measures of executive resource depletion. The fact that the results support our predictions across these varying manipulations and measures suggests that the emotion regulation hypothesis provides the most parsimonious explanation for all four studies.

The current work is consistent with an integrated process model of stereotype threat that Schmader et al. (2008) have recently proposed.

The studies presented here suggest that emotion, and its cognitive collaborator emotion regulation, plays a key role in the experience of stereotype threat. When confronted by negative stereotypes, targets may not only become stressed and agitated but also motivated to regulate these negative emotions. Our research suggests that response-focused coping in the form of emotional suppression can be costly and counterproductive.

Quality of the study - Reporting

Is the context of the study adequately described?

Consider your answer to questions: Why was this study done at this point in time, in those contexts and with those people or institutions? (Section B question 2) Was the study informed by or linked to an existing body of empirical and/or theoretical research? (Section B question 3) Which of the following groups were consulted in working out the aims to be addressed in the study? (Section B question 4) Do the authors report how the study was funded? (Section B question 5) When was the study carried out? (Section B question 6)

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

Are the aims of the study clearly reported?

Consider your answer to questions: What are the broad aims of the study? (Section B question 1) What are the study research questions and/or hypotheses? (Section C question 10)

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

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

Consider your answer to all questions in Methods on 'Sampling Strategy', 'Recruitment and Consent', and 'Actual Sample'.

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

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

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

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

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

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

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

Is the study replicable from this report?

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

Do the authors avoid selective reporting bias?

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

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

Quality of the study - Methods and data

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

Consider consent, funding, privacy, etc.

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

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

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

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

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

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

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

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

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

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

How generalisable are the study results?

☐ Details

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.

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, *137*(4), 691–705. https://doi.org/10.1037/a0013834

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*.