Short Report

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
2020, Vol. 31(7) 865–872

DOI: 10.1177/0956797620923587
www.psychologicalscience.org/PS
SAGE

Can Short Psychological Interventions Affect Educational Performance? Revisiting the Effect of Self-Affirmation Interventions

Marta Serra-Garcia[1], **Karsten T. Hansen**[1], **and Uri Gneezy**[1,2]
[1]Rady School of Management, University of California San Diego, and [2]Center for Research in Experimental Economics and Political Decision Making (CREED), University of Amsterdam

Abstract

Large amounts of resources are spent annually to improve educational achievement and to close the gender gap in sciences with typically very modest effects. In 2010, a 15-min self-affirmation intervention showed a dramatic reduction in this gender gap. We reanalyzed the original data and found several critical problems. First, the self-affirmation hypothesis stated that women's performance would improve. However, the data showed no improvement for women. There was an interaction effect between self-affirmation and gender caused by a negative effect on men's performance. Second, the findings were based on covariate-adjusted interaction effects, which imply that self-affirmation reduced the gender gap only for the small sample of men and women who did not differ in the covariates. Third, specification-curve analyses with more than 1,500 possible specifications showed that less than one quarter yielded significant interaction effects and less than 3% showed significant improvements among women.

Keywords

self-affirmation, gender gap, education, open data

Received 5/13/19; Revision accepted 2/16/20

Education in the United States is far from reaching desired outcomes (Swanson, 2009). The problem is not in the number of policy ideas that are tried; it is in the success of these policies. For example, in a large-scale field experiment to study whether monetary incentives can improve educational outcomes, some students were paid to read books, and others were paid for performance on interim assessments (Fryer, 2011). Despite $9.4 million in spending, the impacts were statistically indistinguishable from zero (see also Fryer, Levitt, List, & Sadoff, 2012; Gneezy, Meier, & Rey-Biel, 2011; Springer et al., 2010).

A particular focus of many programs is to reduce the gender gap in science, technology, engineering, and mathematics (STEM). Given the limited effectiveness and high cost of most interventions, the success of a very simple psychological manipulation tested by Miyake et al. (2010) is impressive. In the intervention, called *values affirmation*, 399 students in a college-level introductory physics class wrote about either their most important values (treatment condition) or their least important values (control condition), twice for 15 min within the first 4 weeks of the 15-week course.

Miyake et al. (2010) reported that the intervention reduced the gender gap in physics exams at the 1% significance level. Whereas studies using monetary incentives to improve performance found average effects ranging from −0.031 to 0.079 standard deviations (Fryer, 2011), the effects in the Miyake et al. study ranged from 0.12 to 0.19 standard deviations. Similar strong results of the self-affirmation intervention have been found in schools (Cohen, Garcia, Apfel, & Master, 2006), even in 2-year and 7- to 9-year follow-up studies

Corresponding Author:
Marta Serra-Garcia, University of California San Diego, Rady School of Management, 9500 Gilman Dr., No. 0553, La Jolla, CA 92093
E-mail: mserragarcia@ucsd.edu

(Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009; Goyer et al., 2017).

Miyake et al. (2010) conducted a widely cited study that has had a substantial impact on the field of STEM education (Hanselman, Rozek, Grigg, & Borman, 2017; Nisbett, 2010) and the broader use of self-affirmation interventions. However, other studies (Borman, 2012; Bratter, Rowley, & Chukhray, 2016; Dee, 2015; de Jong, Jellesma, Koomen, & de Jong, 2016; Hanselman et al., 2017; Hoffman & Kurtz-Costes, 2019; Lauer et al., 2013) have failed to replicate the findings.

We reanalyzed Miyake et al.'s (2010) data, examining it at a conceptual and a statistical level. At the conceptual level, values-affirmation theory starts with the common stereotype that men are better at math and science than are women. This may generate increased academic pressure for women who subscribe to this stereotype and possibly negatively affect their academic performance. To offset this, the theory proposes that a values-affirmation task, in which people reflect on self-defining values, can buffer people against such psychological threat. The resulting hypothesis is that (some) women will improve as a result of the intervention. This theory has no prediction for male students. However, Miyake et al.'s original data showed that the observed reduction in the gender gap was a result of the interaction effect of the intervention on the performance of men and women. We examined the results for male and female students separately.

From a statistical perspective, the analysis by Miyake et al. (2010) was based on covariate-adjusted means. The interaction effect of values affirmation thus must be interpreted as conditional on a given level of prior performance and stereotype endorsement (Cochran, 1957; Miller & Chapman, 2001). However, because prior performance and stereotype endorsement on average differed substantially for men and women, the reported effect was restricted to a small portion (28%) of the sample: those women with the same performance and stereotype endorsement as men.

Miyake et al. (2010) reported the results of one empirical model specification. Using specification-curve analyses (Simonsohn, Simmons, & Nelson, 2015; see also Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016), we examined the robustness of the effect of values affirmation on the gender gap and on women's performance.

Finally, Miyake et al. (2010) argued that results were consistent with theoretical accounts of stereotype threat because the intervention improved the performance of women displaying high stereotype endorsement. To show this, they compared women whose stereotype endorsement was either 0.75 standard deviations higher than the mean (high endorsement) or 0.75 standard deviations lower than the mean (low endorsement). We reexamined these heterogeneous effects considering the responses of all women along the 5-point scale measuring stereotype endorsement.

We also ran a small replication of the Miyake et al. (2010) study in a physics class at the University of California San Diego because replication is an important part of social science (Camerer et al., 2016; Open Science Collaboration, 2012, 2015; Simmons, Nelson, & Simonsohn, 2011; Simonsohn, 2016), and we report the suggestive results of our study in the Supplemental Material available online.

Method

The self-affirmation intervention worked as follows. Students in the treatment group wrote about personally important values (such as friends and family). Students in the control group selected their least important values from the same list and wrote why these values might be important to other people. Thus, all students wrote about values and their importance, but the exercise was self-relevant only for the affirmation group (the students in the treatment group). In the study by Miyake et al. (2010), this 15-min writing exercise was conducted during class once in Week 1 and once in an online homework assignment during Week 4 of the semester of an introductory physics class in college. This brief exercise was not related to the subject matter of the course.

According to Miyake et al. (2010),

> values affirmation, in which people reflect on self-defining values, can buffer people against . . . psychological threat. When they affirm their core values in a threatening environment, people reestablish a perception of personal integrity and worth, which in turn can provide them with the internal resources needed for coping effectively. (p. 1235)

The theory does not predict a negative effect on men but only a positive one for women.

We obtained the raw data from the original study by Miyake et al. (2010) to better understand their results. These data included three continuous measures of performance: mean exam scores, scores on a standardized physics test (the Force and Motion Concept Evaluation, or FMCE), and final course scores, based substantially (75%) on the exam scores.

Results

Using the original data by Miyake et al. (2010), we found that values affirmation had no significant effect

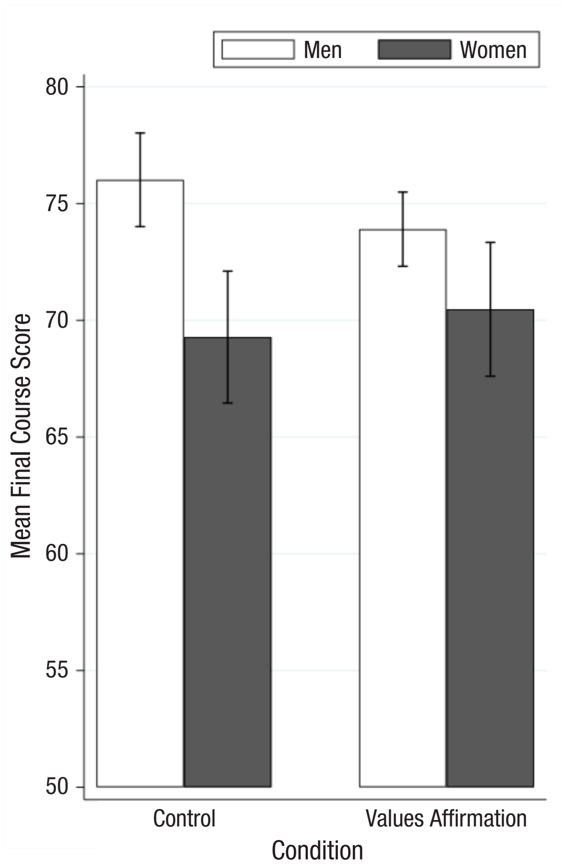

Fig. 1. Mean final course score for men and women in each condition of the Miyake et al. (2010) study. Error bars indicate 95% confidence intervals.

on average female performance (covariate unadjusted). If anything, it had a negative effect on male performance in some dimensions. Covariate-unadjusted means are reported in Table S1 in the Supplemental Material of the Miyake et al. (2010) study.

Values affirmation had no significant effect on the mean exam score of women (Cohen's $d = 0.19$), $t(114) = 1.00$, $p = .318$. However, the performance of male students was significantly lower in the values-affirmation condition relative to the control condition (Cohen's $d = -0.25$), $t(281) = -2.06$, $p = .040$. The interaction effect was significant, $\beta = 0.43$, $t(395) = 1.97$, $p = .050$, but it was driven by a drop in male performance.

Values affirmation did not have a significant effect on male or female scores on a standardized physics test (FMCE score) at the end of the semester. The scores of female students did not differ significantly between the values-affirmation condition and the control condition (Cohen's $d = 0.31$), $t(94) = 1.28$, $p = .21$. Male students did not obtain significantly different scores between conditions either (Cohen's $d = -0.09$), $t(210) = -0.54$, $p = .587$. The interaction effect between values affirmation and gender was not significant, $\beta = 0.34$, $t(304) = 1.40$, $p = .164$.

Figure 1 presents the results obtained by Miyake et al. (2010) for the final course scores. The final course scores obtained by women and men were not affected by the values-affirmation intervention, and the interaction effect was not significant either. All details of the regression analysis are reported in the Supplemental Material.

The Interpretation of Covariate-Adjusted Effects

These results contrast with the covariate-adjusted results reported by Miyake et al. (2010)—that there was a significant positive effect of the values-affirmation intervention on the performance of women using covariate-adjusted means. The adjustment was based on two covariates. The first was prior performance, measured by the SAT score for the mean exam score and the course score and by the beginning-of-semester FMCE score for the FMCE score at the end of the semester. The second covariate was stereotype endorsement, measured by students' agreement with the statement, "According to my own personal beliefs, I expect men to generally do better in physics than women," on a 5-point scale ranging from *strongly disagree* to *strongly agree*. The authors argued that it is critical to assess the effects of the self-affirmation intervention "controlling for prior relevant performance" (Miyake et al., 2010, Supplemental Material, p. 13) and that the interaction of background variables with gender, condition, and stereotype endorsement should be included. They pooled male and female students and examined the effect of the self-affirmation intervention, including variables for gender, stereotype endorsement, and prior performance in their analysis, as well as the interaction terms between them.

The interpretation of covariate-adjusted effects is that there was a reduction in the gender gap for a population of men and women who have the same SAT scores (and level of stereotype endorsement). We provide a formal interpretation of covariate-adjusted effects in the Supplemental Material. Is the covariate-adjusted effect the one that we are interested in? It could be if the distributions of SAT scores and stereotype endorsement for men and women had similar means. However, there was a significant difference in the prior performance of men and women—for SAT scores, $t(397) = 2.62$, $p = .01$; for beginning-of-semester FMCE scores, $t(306) = 4.80$, $p < .01$—and in stereotype endorsement, $\chi^2(4, N = 399) = 41.64$, $p < .01$. Hence, the effect that was estimated with covariate adjustment was relevant only for the subset of women who had the same prior performance and stereotype endorsement as men. In the Miyake et al. (2010) study, this was 56% of the sample, considering only SAT or ACT scores. Including stereotype endorsement, we found that only 28% of the sample featured male and female students with the same SAT scores and stereotype endorsement.

The relevant question is whether values affirmation had a significantly positive effect on average female performance. We examined this effect as part of the specification-curve analysis described next.

Evaluating Analytical Approaches: Specification-Curve Analysis

As we discussed above, adding covariates such as SAT scores and their interactions changes the interpretation of the coefficients and also invalidates standard linear model analysis. Researchers may be interested in understanding how robust the effect of self-affirmation interventions is to the inclusion of different covariates in the regression model, as well as other decisions regarding the data analysis (e.g., Simonsohn et al., 2015; Steegen et al., 2016). To investigate this, we conducted a specification-curve analysis (Simonsohn et al., 2015).

The regression model used by Miyake et al. (2010) included 11 independent variables. In addition to including gender (F_i), affirmation condition (Z_i), and prior performance (S_i), the model included stereotype threat, which we denote as T_i. The specification included two-way and three-way interaction effects of these variables and was as follows:

$$\begin{aligned} Y_i = \beta + \beta_Z \times Z_i + \beta_F \times F_i + \beta_{FZ} \times F_i \times Z_i + \beta_S \times S_i \\ + \beta_{SZ} \times S_i \times Z_i + \beta_{SF} \times S_i \times F_i + \beta_T \times T_i + \beta_{TZ} \\ \times T_i \times Z_i + \beta_{TF} \times T_i \times F_i + \beta_{TS} \times T_i \times S_i + \beta_{TZF} \\ \times T_i \times Z_i \times F_i + \varepsilon_i \end{aligned}$$

Miyake et al. (2010) focused on the average exam score as their main outcome of interest but also included the final course score and score on the FMCE at the end of the semester in some of the analyses. We investigated how the gender gap in academic performance (Values Affirmation × Gender interaction) changed given the following alternatives: (a) using the three different dependent variables mentioned above, all of which measure academic performance; (b) including stereotype threat as a covariate, with or without its interaction with other covariates; (c) using different definitions of stereotype threat (as a continuous variable or a dummy for those above the median); (d) including prior performance, with or without its interaction with other covariates; (e) using different variables and definitions of prior performance (SAT/ACT score or FMCE score at the beginning of the semester, continuous or as a dummy for those performing above the median); (f) allowing for different exclusions of missing observations; (g) and estimating robust standard errors. These decisions yielded a total of 1,566 unique estimated interaction effects of values affirmation and gender. Further details of the analysis and results are provided in the Supplemental Material. Because the relevant question was whether the values affirmation had a significantly positive effect on average female performance, we also conducted the specification-curve analysis for the effect of values affirmation on female students.

Figure 2a plots the coefficient and confidence interval of the interaction of values affirmation and gender at the sample average. The dependent variable was standardized so the coefficient can be interpreted in standard deviations. Figure 2b plots the coefficient and confidence interval of the effect of values affirmation on female students. Each panel also shows the distribution of effects for each of the researchers' decisions, described on the left.

We replicated the t statistic for the interaction effect found by Miyake et al. (2010), 3.08, and indicate where it lies on the specification curve. It was the 15th highest out of 1,566 specifications. The interaction effect was not statistically significant ($p > .05$) in 1,205 specifications. That is, in 77% of specifications, the effect of values affirmation was not different between female and male students. As shown in Figure 2, a key decision in obtaining a significant interaction effect was to include the three-way interaction among values affirmation, gender, and stereotype-threat endorsement. When it was not included, 91% of specifications (1,000 out of 1,098) were not significant. By contrast, of the 361 significant interaction effects, 73% included the three-way interaction. This indicates not only that covariate adjustment was needed but also that a three-way interaction term was needed to find a significant reduction of the gender gap in academic performance.

Within the group of female students, the treatment was assigned randomly, and women in the treatment and control conditions did not vary in their prior performance—for SAT scores by condition, $t(114) = -0.79$, $p = .4329$; for beginning-of-semester FMCE scores by condition, $t(94) = -0.65$, $p = .5187$—or stereotype endorsement, $\chi^2(4, N = 116) = 2.46$, $p = .653$. Hence, covariate adjustment could be used to reduce variance (Miller & Chapman, 2001) without reducing the effect of values affirmation to a specific group of students. In the data shared with us by Miyake et al. (2010), SAT/ACT grades were standardized for men and women, and with the data available, we could not recode them and use them within the sample of women. Stereotype

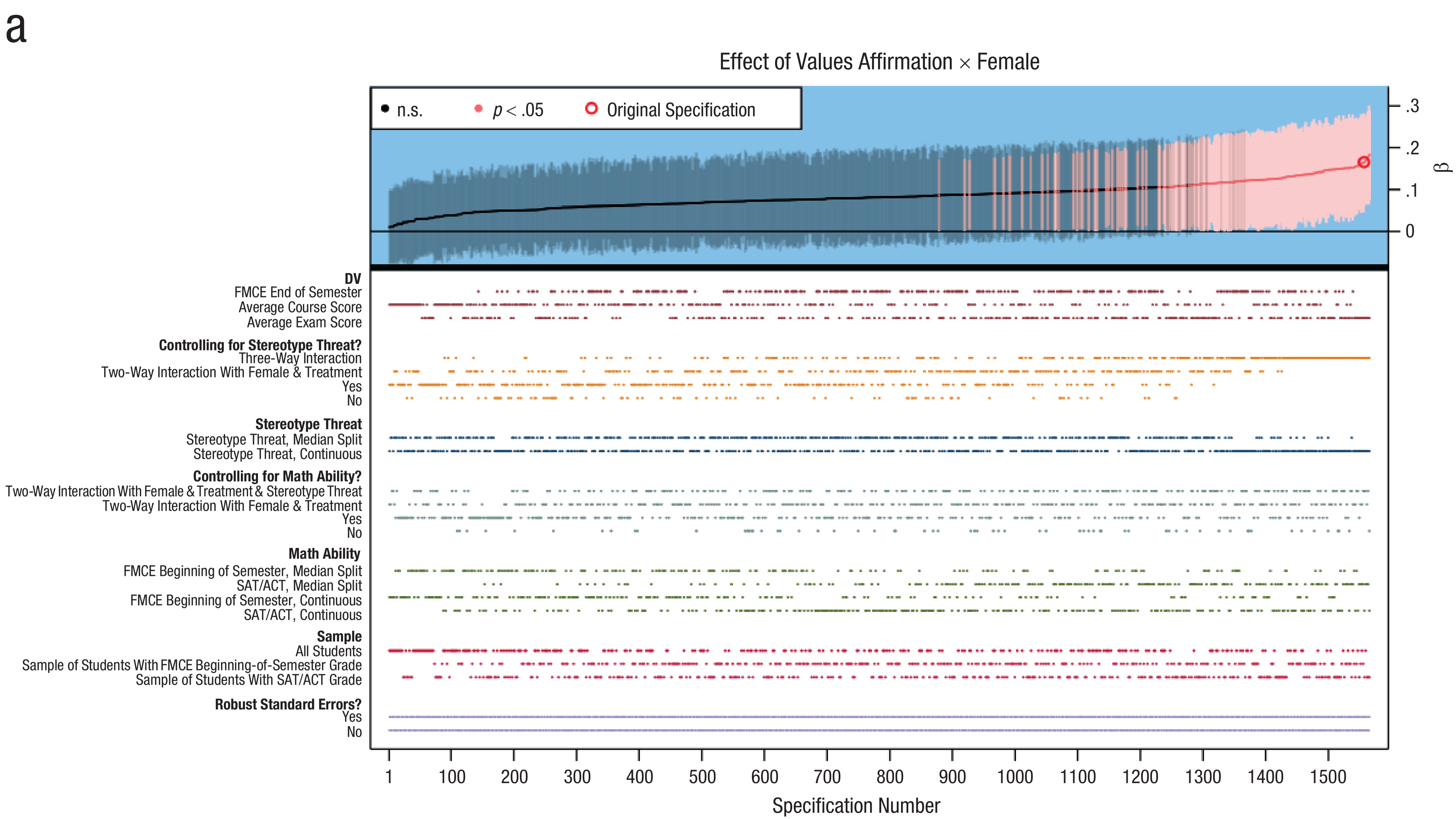

Fig. 2. *(continued on next page)*

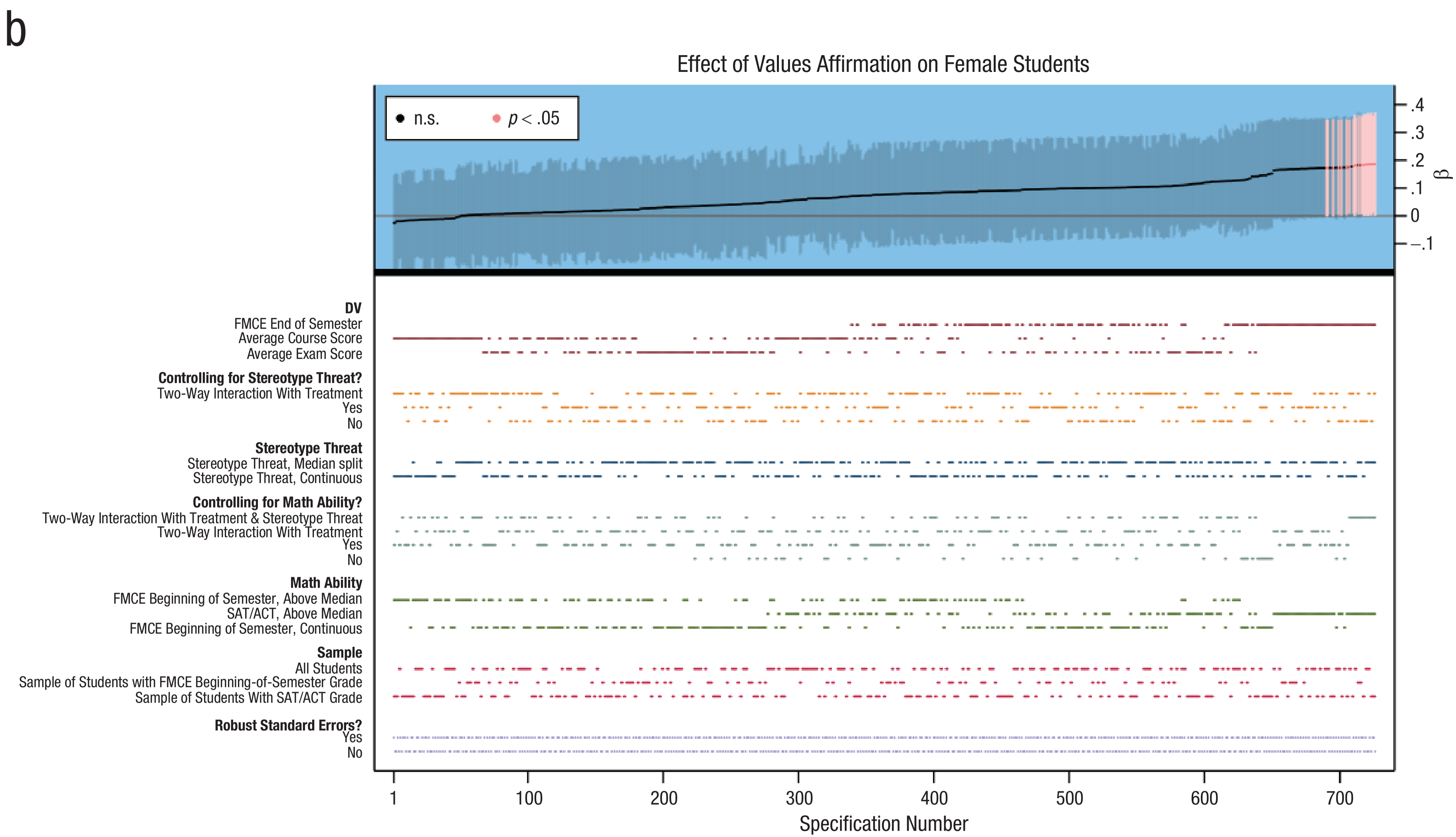

Fig. 2. Specification-curve analysis for (a) the effect of values affirmation on the gender gap and (b) the effect of values affirmation on female students' performance. Standardized coefficients, along with 95% confidence intervals, are shown in the top half of each panel for each of the researchers' decisions, which are indicated by the dots in the bottom half of each panel. Performance was always measured on the standardized dependent variable (DV). In (a), the interaction effect found by Miyake et al. (2010) is indicated by the red circle. To conduct the analysis, we could not use the SAT/ACT grade because it was standardized for male and female students in the original data, and its raw value could not be recovered. FMCE = Force and Motion Concept Evaluation.

endorsement and FMCE scores at the beginning of the semester were sample centered. They could be recoded, and we used these as covariates within the sample of women. The specification curve is shown in Figure 2b. The effect of values affirmation on female students was not significant in 97.0% of specifications.

Heterogeneity of Values-Affirmation Effects by Stereotype Endorsement

Miyake et al. (2010) argued that "values affirmation was particularly beneficial for women who tended to endorse the gender stereotype" (p. 1236). They split the sample between high and low stereotype endorsement on the basis of whether the student was, respectively, 0.75 standard deviations above or below the mean. We examined this finding in further detail to understand the role of this particular sample split. Stereotype endorsement was based on agreement with the statement, "According to my own personal beliefs, I expect men to generally do better in physics than women," with responses ranging from *strongly disagree* to *strongly agree* on a 5-point scale. Out of 96 female students without missing information, 7 (4 treatment, 3 control) agreed with the stereotype, 10 (4 treatment, 6 control) neither agreed nor disagreed, 24 disagreed (14 treatment, 10 control), and 55 (33 treatment, 22 control) strongly disagreed. Values affirmation had a positive and significant effect on the exam scores of women who agreed with the stereotype (4 women) and women who neither agreed nor disagreed (a total of 8 women). Surprisingly, for women who strongly disagreed with the statement, we observed negative effects of values affirmation on their exam score (33 women).

Using a regression analysis of performance on the female sample and controlling for stereotype endorsement and prior performance, we found that the interaction between values affirmation and stereotype endorsement was significantly positive. Yet this indicates only that the effect of values affirmation was more positive on the students with higher stereotype endorsement. It should not be interpreted as being a positive effect for everyone and more strongly positive for women. As described above, the effect was negative for a large group of students. Detailed results are presented in the Supplemental Material.

Discussion

Understanding how to improve academic performance, in particular for struggling students, is an important challenge for social scientists and policymakers. Miyake et al. (2010) offered a remarkably strong and cost-effective way of doing that. Our investigation was motivated by a desire to better understand the psychological instrument in the Miyake et al. study and use it in large-scale interventions. A closer look at Miyake et al.'s data revealed two critical problems. The self-affirmation hypothesis predicted an effect only on women. However, their conclusions were based on the significant interaction effect between the self-affirmation intervention and gender. This effect was driven by a reduction in the performance of men and not by an improvement in the performance of women. Further, the conclusions from the original study were highly sensitive to the empirical specification. Miyake et al. wrote that "values affirmation is a promising intervention that can help reduce the gender achievement gap in physics" (p. 1237). The statistical analysis presented here suggests that this conclusion was not supported by the data. Taking the evidence together, we found that the results were supportive of a null effect. It is important to carefully understand the effects of such interventions because promoting ineffective interventions is costly in terms of resource allocation and negatively affects the success of policymakers' attempts to reduce the gender achievement gap.

Transparency

Action Editor: Brent W. Roberts
Editor: D. Stephen Lindsay
Author Contributions

U. Gneezy and M. Serra-Garcia developed the study concept. K. T. Hansen analyzed the regression models in the Miyake et al. (2010) study and interpreted the covariate-adjusted effects. M. Serra-Garcia reanalyzed the data in the Miyake et al. study. All the authors drafted and revised the manuscript and approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

All data and STATA code necessary to reproduce the analyses have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/gzq9k. The design and analysis plans were not preregistered. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620923587. This article has received the badge for Open Data. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iD

Marta Serra-Garcia https://orcid.org/0000-0002-5875-4986

Acknowledgments

We thank Tiffany Ito for sharing the materials and data, and we thank Anna Dreber, Leif Nelson, and Johannes Haushofer for valuable comments. This project was conducted under

University of California San Diego Institutional Review Board Approval No. 130701XX. Data were obtained from the authors of the original study (Miyake et al., 2010).

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620923587

References

Borman, G. (2012). *Examination of a self-affirmation intervention in St. Paul Public Schools* (Vol. 12). Washington, DC: The Council of the Great City Schools.

Bratter, J., Rowley, K. J., & Chukhray, I. (2016). Does a self-affirmation intervention reduce stereotype threat in Black and Hispanic high schools? *Race and Social Problems*, *8*, 340–356.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., . . . Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*, 1433–1436.

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics*, *13*, 261–281.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science*, *313*, 1307–1310.

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, *324*, 400–403.

Dee, T. S. (2015). Social identity and achievement gaps: Evidence from an affirmation intervention. *Journal of Research on Educational Effectiveness*, *8*, 149–168.

de Jong, E. M., Jellesma, F. C., Koomen, H. M. Y., & de Jong, P. F. (2016). A values-affirmation intervention does not benefit negatively stereotyped immigrant students in the Netherlands. *Frontiers in Psychology*, *7*, Article 691. doi:10.3389/fpsyg.2016.00691

Fryer, R. G., Jr. (2011). Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics*, *126*(4), 1755–1798.

Fryer, R. G., Jr., Levitt, S. D., List, J., & Sadoff, S. (2012). *Enhancing the efficacy of teacher incentives through loss aversion: A field experiment* (NBER Working Paper No. 18237). Cambridge, MA: National Bureau of Economic Research.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, *25*, 191–209.

Goyer, J. P., Garcia, J., Purdie-Vaughns, V., Binning, K. R., Cook, J. E., Reeves, S. L., . . . Cohen, G. L. (2017). Self-affirmation facilitates minority middle schoolers' progress along college trajectories. *Proceedings of the National Academy of Sciences, USA*, *114*, 7594–7599.

Hanselman, P., Rozek, C. S., Grigg, J., & Borman, G. D. (2017). New evidence on self-affirmation effects and theorized sources of heterogeneity from large-scale replications. *Journal of Educational Psychology*, *109*, 405–424.

Hoffman, A. J., & Kurtz-Costes, B. (2019). Promoting science motivation in American Indian middle school students: An intervention. *Educational Psychology*, *39*, 448–469.

Lauer, S., Momsen, J., Offerdahl, E., Kryjevskaia, M., Christensen, W., & Montplaisir, L. (2013). Stereotyped: Investigating gender in introductory science courses. *CBE-Life Sciences Education*, *12*, 30–38.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*(1), 40–48.

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, *330*, 1234–1237.

Nisbett, R. E. (2010). Think big, bigger . . . and smaller. *Educational Leadership*, *68*(3), 10–15.

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*, 657–660.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article 4716. doi:10.1126/science.aac4716

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.

Simonsohn, U. (2016). Each reader decides if a replication counts. *Psychological Science*, *27*, 1410–1412.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification curve: Descriptive and inferential statistics on all reasonable specifications. *SSRN*. doi:10.2139/ssrn.2694998

Springer, M. G., Ballou, D., Hamilton, L., Le, V.-N., Lockwood, J. R., McCaffrey, D. F., . . . Stecher, B. M. (2010). *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching*. Retrieved from https://www.rand.org/content/dam/rand/pubs/reprints/2010/RAND_RP1416.pdf

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*, 702–712.

Swanson, C. B. (2009). *Cities in crisis 2009: Closing the graduation gap*. Bethesda, MD: Editorial Projects in Education.