# Replication of a Research Claim from Seaton et al. (2010), from the American Educational Research Journal

Project ID: Seaton_AmEduResJourn_2010_Blxd - Nast/Field - Data Analytic Replication – 3053

OSF Project Link: https://osf.io/mu4rs/

Preregistration Link: https://osf.io/cqm5f/

Replication Team: Carolin Nast (Utrecht University); James G. Field (West Virginia University)

Research Scientist: Melissa Kline

Action Editor: John Lloyd

Final Report

November 19, 2020

**Replication of a research claim from Seaton et al. (2010), from the Journal of Environmental Psychology**

## Claim Summary

The claim selected for replication from Seaton et al. (2010) is that larger BFLPE (big-fish-little-pond) effects were associated with students who used memorization to a greater extent. The original study utilizes the Programme for International Student Assessment (PISA) by the OECD from 2003, which had a focus on students' mathematical abilities. Participants were 15-year-old students from 41 countries (N = 276,165). PISA assesses students' mathematics, literacy, and science, but PISA also contains a vast amount of additional information about students' families, home backgrounds, and SES.

Students in high-ability schools who used the memorization technique to a greater extent suffered a larger decline in mathematics self-concept than those who used the technique to a lesser extent. Although all students who used memorization had lower mathematics self-concepts if they attended high-ability schools than students of similar memorization usage who attended average- or low-ability schools, the drop in mathematics self-concept was more pronounced for students who used memorization to a greater extent. This reflects the following statement from the paper's abstract: "Statistically significant moderating effects emerged in both areas; however, in relation to the large sample (N = 265,180), many were considered small." The claim was tested by regressing mathematical self-concept on the following predictor variables: individual mathematics ability (linear and quadratic), school-average mathematics ability, the moderator of interest - use of memorization studying, and the interaction of school-average mathematics ability with memorization. Due to the large number of tests of statistical significance being conducted, the significance level was set at $p < .001$. Effect size for the group-level construct of BFLPE was calculated using the equation shown on p. 406; effect sizes that approached 0.10 or greater were considered of interest. Use of memorization moderated the effect of school-average ability on mathematical self-concept (-.089, $p < 0.001$; see Table 3), with an effect size of -0.1 57 (see Table 4).

We followed analytic procedures described in the original study to conduct a multilevel analysis using a newer PISA wave, namely PISA 2012.

**Deviations from the Original Study**

With regard to data collection, the only deviation, to the best of our knowledge, there is only difference between the original and replication studies. Specifically two constructs, (Cooperative Orientation and Competitive Orientation) were officially deleted from the PISA student questionnaire after 2003 and, thus, did not appear in the PISA 2012 data collection effort. Therefore, a replication attempt would need to be followed through without these two constructs. This could have implications on the replication attempt, as these independent variables were significant, however, their standardized effect sizes are considered very small (< 0.1 ). Furthermore, given the fact that these constructs were deleted due to potential invalidity, the results from the original study concerning these independent variables might also be not totally understood/valid. Consequently, it is expected that the non-availability of these independent variables for the replication attempt is minor due to potential invalid previous results.

**Deviations from the Preregistration**

Two deviations must be reported. First, we describe an error in the analytic script used to reanalyze the focal claim and the steps taken to produce a better replication attempt. Second, we describe how the ordering in which missing data were removed was revised to better match what was described in the original article.

*Error in analytic script*. On page 405 of the original article (Seaton et al., 2010), the original authors specified the fixed components of each multilevel regression model. Specifically, they wrote that

> "… the fixed components were individual ability (both linear and quadratic), the specific moderator, school-average ability, and the cross-product of the moderator and school-average ability. All models tested had a three-level structure: Individual students were at Level 1, schools were at Level 2, and country was at Level 3."

The Data Analyst followed these and other procedures outlined in the Methods section of the original article to produce the multilevel regression models needed to replicate the focal analysis. Unfortunately, the original authors did not specify the random components of each multilevel model and, to the best of our knowledge, the Methods section of the original article indicates that random effects for school and country only needed to be specified. However, an examination of Table 3 (see p. 411) indicates that the random effects components included Level 3 (country intercept, linear ability, quadratic ability, school-average ability, and moderator), Level 2 (school intercept, linear ability, quadratic ability, and moderator), and Level 1 (individual intercept) variables. Thus, the original models (i.e., the ones used to produce the original replication result) were not specified correctly – some of the random effects were not included. This discrepancy was brought to the SCORE team's attention on Tuesday, October 27, 2020. On Wednesday, November 18, 2020, the Data Analyst was given permission by the SCORE team to revise the analytic script to include the missing random effects and update the replication analysis result. Both the SCORE team and the Data Analyst agree that the revised analytic script offers a better test of the focal analysis claim and, thus, should be used as evidence for the SCORE project, rather than the preregistered analysis.

Taken together, the Data Analyst revised the analytic script multilevel regression models to match the output (i.e., fixed and random effects) reported in Table 3 of the original article (see https://osf.io/qrcpe/?view_only=f8c113e4a21b45bca06c091a5db3a411). However, it is not clear why a Level 1 variable (i.e., the individual student) was included in the random effects component of the model because there is only one observation per individual in the dataset. It would make sense to include a random effect for the individual if multiple observations were recorded for each student, but an examination of the replication dataset indicates that only one observation per student was recorded. As such, we are not sure how a random effect can be captured for the individual intercept when there is only one observation per "unit" at Level 1. Thus, the analytic script was revised to include all random effects reported in Table 3 of the original article (except for the random effect at Level 1 because is only one observation per student).

*Missing data*. In the original analytic script, observations were removed in the following order.

(1) Observations nested in schools with 10 or fewer students were removed
(2) Observations with missing data in the focal variables were removed (i.e., math self-concept [DV], plausible values 1-5 [IVs], and memorization [moderator]

However, a closer inspection of the Methods section indicates that the ordering in which observations are removed should have been reversed (see page 403 of the original article; https://osf.io/qrcpe/?view_only=f8c113e4a21b45bca06c091a5db3a411). As such, in the revised analytic script, observations are removed in the following order:

(1) Observations with missing data in the focal variables were removed (i.e., math self-concept [DV], plausible values 1-5 [IVs], and memorization [moderator]
(2) Observations nested in schools with 10 or fewer students were removed

**Replication Criteria**

A successful replication attempt is observed if the replication analysis produces a statistically significant effect (alpha = .05, two tailed) in the same pattern as the original study for the focal hypothesis test (**H***). The hypothesis to be tested is: The interaction of the use of memorization and school-average ability will be negative in its association with mathematical self-concept. As such, the inference criteria that will be used to assess whether or not the focal claim replicates will be (1 ) the direction of the final parameter estimate (needs to be negative) and (2) the magnitude of the final *p*-value (needs to be less than less than .001 [see page 405 of Seaton et al., 201 0]. Both conditions need to be met in order for the focal claim to replicate.

**Replication Result**

First, we describe the original replication result, which was produced on September 14, 2020 and before the Data Analyst noticed an error in the analytic script used to replicate the focal claim result (see earlier description). Following the description of the original replication result, a description of the replication result that was produced using a revised analytic script is provided.

*Description of original replication result.* The replication included 150,331 participants nested within 16,007 and 65 schools and countries, respectively in the analytic sample to reach the requirements for Stage 1 data collection. The primary unit of analysis is the school. An estimate of the minimum viable sample size for the data analytic replication is 972. Furthermore, the Stage 1 and Stage 2 required sample sizes would be 4,711 and 10,619, respectively. As such, the replication study's sample size is large enough to achieve the desired level of statistical power. Our multilevel modeling analysis revealed that the observed coefficient for the focal moderating effect was 0.001370106 (SE = 0.003331732) with a corresponding *p*-value of .84. Thus, this replication of the claim was not successful according to the SCORE criteria.

*Description of revised replication result.* The replication included 68,088 students nested in 5,298 and 59 countries respectively. to reach the requirements for Stage 1 data collection. The primary unit of analysis is the school. An estimate of the minimum viable sample size for the data analytic replication is 972. Furthermore, the Stage 1 and Stage 2 required sample sizes would be 4,711 and 10,619, respectively. As such, the replication study's sample size is large enough to achieve the desired level of statistical power. Our multilevel modeling analysis revealed that the observed coefficient for the focal moderating effect was -0.02609457 (SE = 0.008943238) with a corresponding *p*-value of 3.641247e-10. Thus, this replication of the claim was successful according to the SCORE criteria – the replication produced a statistically significant result ($p < .001$) that corresponded to an effect in the same pattern compared to the original study. In other words, the replication result supports the notion that students in high-ability schools who used the memorization technique to a greater extent suffered a larger decline in mathematics self-concept than those who used the technique to a lesser extent.

## Methods & Materials

The following materials are publicly available on the OSF site (see https://osf.io/mu4rs/):

- The **SCORE report** that describes the observations and results of the replication attempt. Filename:
    - SCORE Report_Seaton_AmEduResJourn_2010_Blxd - Nast and Field - 3053_withCorrection

- The completed and approved **preregistration form** that describes all aspects of the replication attempt, downloaded from the OSF, and provided as a PDF file. Filename:
    - Seaton_AmEduResJourn_2010_Blxd_3053 (Nast_Field) Preregistration.pdf
    - Note that the preregistration form describes the analysis plan for the original replication result, which is not used as evidence for SCORE. The results described the "*Description of revised replication result*" is being used as the evidence for SCORE, rather than the preregistered analysis, as the Data Analyst and the SCORE team are in agreement that this should be the focal replication result.
- The **power analysis** materials, provided as a zip file. Filename:
    - POWER_Seaton_AmEduResJourn_2010_Blxd.zip
        - This zip folder contains three separate file, which together can be used to replicate the required sample size estimation for the replication attempt. The filenames are:
            - ss_Seaton_AmEduResJourn_2010_Blxd_dar.md
            - Seaton_AmEduResJourn_2010_Blxd_power_vars.tsv
            - ss_Seaton_AmEduResJourn_2010_Blxd_dar.Rmd
- The **raw PISA 2012 dataset** with the full set of responses from individual students, downloaded directly from http://www.oecd.org/pisa/data/pisa2012database-downloadabledata.htm. This file is provided as .txt file. Filename:
    - INT_STU12_DEC03.txt
- The **source for creating indices** CSTRAT, MEMOR and ELAB, provided as a PDF file. Filename:
    - OECD Education Working Papers No. 130.pdf
- The relevant chapter from the **PISA Data Analysis Manual** (2nd SPSS Edition), provided as a PDF. Filename:
    - OECD Plausible Values.pdf
- The **OECD Technical report (2005)** concerning PISA 2003, as a PDF file. Filename:
    - OECD Technical report, (2005).pdf
- The **OECD Technical report (2014)** concerning PISA 2012, provided as a PDF file. Filename:

    - OECD Technical report, (2014).pdf
- The official **codebook** for PISA 2012 main study student questionnaire, provided as a PDF file. Filename:
    - PISA12_codebook_student questionnaire.pdf
- The official **questionnaire** for PISA 2012 main study student questionnaire, provided as a PDF file. Filename:
    - PISA12_Student Questionnaire_FormA_ENG.pdf
- The **replication data set**, provided as a .RDS file. Filename:
    - PISA2012.replication.RDS
- The SAS control file used to process the data, provided as a .SAS file and downloaded directly from http://www.oecd.org/pisa/data/pisa2012database-downloadabledata.htm. Filename:
    - PISA2012_SAS_student.sas
- The PISA Data Analysis Manual (SAS Second Edition), provided as a PDF file. Filename:
    - Pisa Data Analysis Manual.pdf
- A **data dictionary** for every variable in the analytic dataset, provided as an Excel file. Filename:
    - PISA Data Dictionary.xlsx
- A **data analysis script,** provided as an R Markdown file. This script transforms the raw data into the analysis data set. Filename:
    - SEATON.code.rmd
- A **data analysis script**, provided as an .R file. This file contains the revised, corrected analytic script that produces the revised replication analysis result (i.e., the result that is being used as evidence for SCORE). Filename:
    - !!CORRECTION!!)_Seaton_AmEduResJourn_2010_Blxd_final.R
    - The entire replication analysis can be replicated by running this data analysis script.
- A **data analysis script**, provided as an .R file. This file contains the erroneous analytic script that produces the original replication analysis result (i.e., the result not being used as evidence for SCORE). Filename:

    - Seaton_AmEduResJourn_2010_Blxd_final.R
- A **data analysis script**, provided as an .R file. On pages 9-10 of the corresponding preregistration form (see https://osf.io/cqm5f/) the Data Analyst discloses that they inadvertently ran a portion of the final analytic script on the *entire* replication data set – not just a 5% random sample of the replication data set, which they were instructed to do by the SCORE team. Consequently, the Data Analyst observed a set of results derived from the set of multilevel modeling regression analyses. This file contains the code that produced these results. Filename:
    - Seaton_AmEduResJourn_2010_Blxd_beta.R

**Replication Dataset Access/License Information**

According to the OECD's terms and conditions (see "Permitted use" in Section 1 (c) at http://www.oecd.org/termsandconditions/, individuals "can extract from, download, copy, adapt, print, distribute, share and embed Data for any purpose, even for commercial use." We are thankful to the OECD for providing open access to the PISA 2012 dataset.

**Data Overlap Disclosure**

This statement confirms that there was no overlap between the original and replication datasets. Specifically, the two datasets were independent.

**Citations**

OECD (2013), PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en

Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-fish-little-pond effect: Generalizability and moderation—Two sides of the same coin. American Educational Research Journal, 47(2), 390-433. doi: 10.3102/0002831209350493