Replication-based learning in Mathematics Education

Ulises Xolocotzin Eligio

Mathematics Education Department, Cinvestav

Implementation and Replication Studies in Mathematics Education Conference

Mexico City, 2025

Introduction

1. Improving the reliability of research in Mathematics Education requires individual and cultural shifts in our community, including the adoption of replication studies.

2. Such shifts require the introduction of replication practices in the apprenticeship of future Mathematics Education researchers.

I. Replication and credibility: Community changes in Psychology

The replication crisis

WORLD VIEW *A personal take on events*

UNIV. MICHIGAN PHOTO SERVICES

The road to fraud starts with a single step

The extensive academic fraud of Diederik Stapel has rocked science. Social psychologist **Jennifer Crocker** *traces the destructive path that cheats follow.*

Retraction of "The influence of mood on attribution," "Affects of the unexpected: When inconsistency feels good (or bad)," "Why people stereotype affects how they stereotype: The differential influence of comprehension goals and self-enhancement goals on stereotyping," "Silence and table manners: When environments activate norms," and "Event accessibility and context effects in causal inference: Judgment of a different order."

Personality and Social Psychology Bulletin
38(10) 1378

DOI: 10.1177/0146167212462821
http://pspb.sagepub.com
SAGE

Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 407–425

0022-3514/11/$12.00 DOI: 10.1037/a0021524

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented for 4 time-reversed effects: precognitive approach to erotic stimuli and precognitive avoidance of negative stimuli; retroactive priming; retroactive habituation; and retroactive facilitation of recall. The mean effect size (*d*) in psi performance across all 9 experiments was 0.22, and all but one of the experiments yielded statistically significant results. The individual-difference variable of stimulus seeking, a component of extraversion, was significantly correlated with psi performance in 5 of the experiments, with participants who scored above the midpoint on a scale of stimulus seeking achieving a mean effect size of 0.43. Skepticism about psi, issues of replication, and theories of psi are also discussed.

Keywords: psi, parapsychology, ESP, precognition, retrocausation

Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 426–432

0022-3514/11/$12.00 DOI: 10.1037/a0022790

Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011)

Eric–Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, and Han L. J. van der Maas
University of Amsterdam

Does psi exist? D. J. Bem (2011) conducted 9 studies with over 1,000 participants in an attempt to demonstrate that future events retroactively affect people's responses. Here we discuss several limitations of Bem's experiments on psi; in particular, we show that the data analysis was partly exploratory and that one-sided *p* values may overstate the statistical evidence against the null hypothesis. We reanalyze Bem's data with a default Bayesian *t* test and show that the evidence for psi is weak to nonexistent. We argue that in order to convince a skeptical audience of a controversial claim, one needs to conduct strictly confirmatory studies and analyze the results with statistical tests that are conservative rather than liberal. We conclude that Bem's *p* values do not indicate evidence in favor of precognition; instead, they indicate that experimental psychologists need to change the way they conduct their experiments and analyze their data.

Keywords: confirmatory experiments, Bayesian hypothesis test, ESP

Replicability and reproductibility

RESEARCH

RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

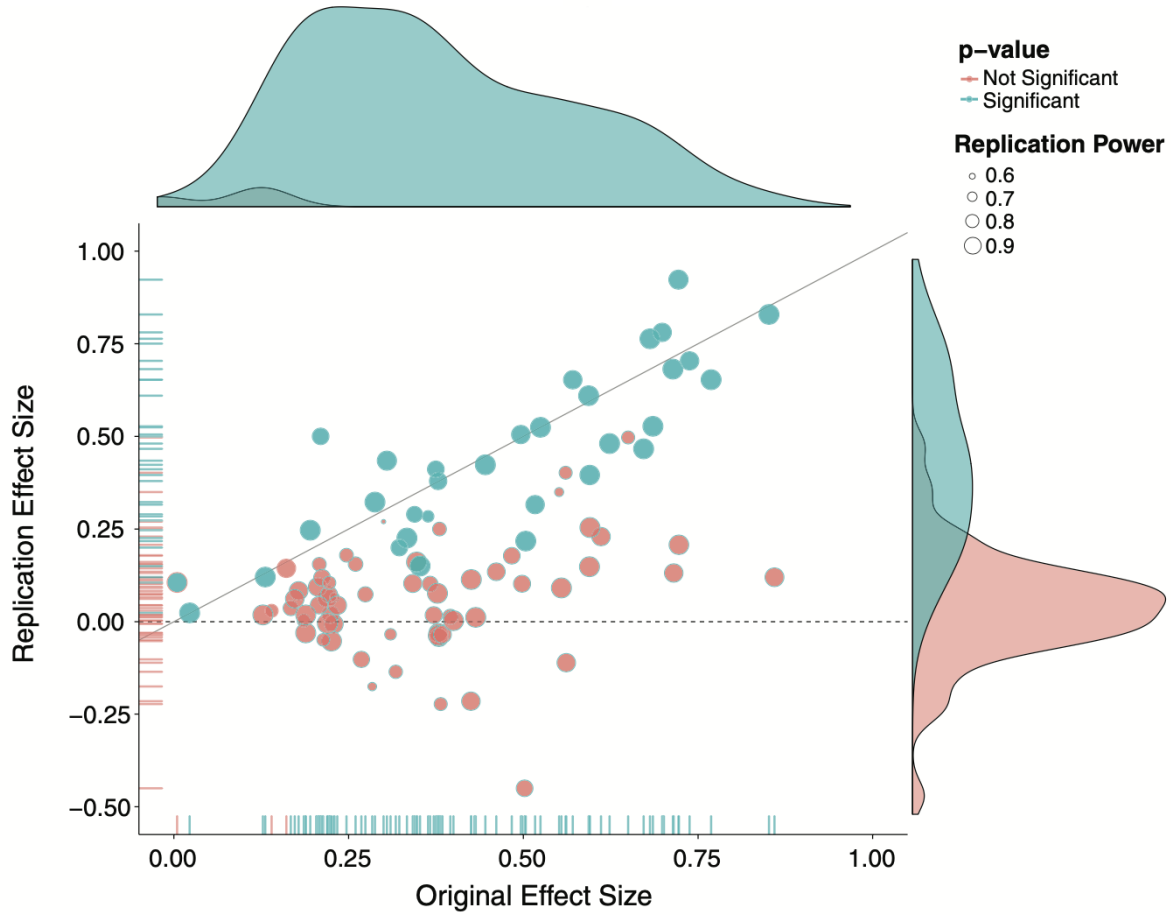

Fig. 3. Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.

Community change

nature human behaviour

PERSPECTIVE

PUBLISHED: 10 JANUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0021

OPEN

A manifesto for reproducible science

Marcus R. Munafò[1,2*], Brian A. Nosek[3,4], Dorothy V. M. Bishop[5], Katherine S. Button[6], Christopher D. Chambers[7], Nathalie Percie du Sert[8], Uri Simonsohn[9], Eric-Jan Wagenmakers[10], Jennifer J. Ware[11] and John P. A. Ioannidis[12,13,14]

Table 1 | A manifesto for reproducible science.

Theme	Proposal	Examples of initiatives/potential solutions (extent of current adoption)	Stakeholder(s)
Methods	Protecting against cognitive biases	All of the initiatives listed below (* to ****) Blinding (**)	J, F
	Improving methodological training	Rigorous training in statistics and research methods for future researchers (*) Rigorous continuing education in statistics and methods for researchers (*)	I, F
	Independent methodological support	Involvement of methodologists in research (**) Independent oversight (*)	F
	Collaboration and team science	Multi-site studies/distributed data collection (*) Team-science consortia (*)	I, F
Reporting and dissemination	Promoting study pre-registration	Registered Reports (*) Open Science Framework (*)	J, F
	Improving the quality of reporting	Use of reporting checklists (**) Protocol checklists (*)	J
	Protecting against conflicts of interest	Disclosure of conflicts of interest (***) Exclusion/containment of financial and non-financial conflicts of interest (*)	J
Reproducibility	Encouraging transparency and open science	Open data, materials, software and so on (* to **) Pre-registration (**** for clinical trials, * for other studies)	J, F, R
Evaluation	Diversifying peer review	Preprints (* in biomedical/behavioural sciences, **** in physical sciences) Pre- and post-publication peer review, for example, Publons, PubMed Commons (*)	J
Incentives	Rewarding open and reproducible practices	Badges (*) Registered Reports (*) Transparency and Openness Promotion guidelines (*) Funding replication studies (*) Open science practices in hiring and promotion (*)	J, I, F

Estimated extent of current adoption: *, <5%; **, 5–30%; ***, 30–60%; ****, >60%. Abbreviations for key stakeholders: J, journals/publishers; F, funders; I, institutions; R, regulators.

communications psychology

PERSPECTIVE

https://doi.org/10.1038/s44271-023-00003-2 OPEN

The replication crisis has led to positive structural, procedural, and community changes

Max Korbmacher[1,2,3], Flavio Azevedo[4,35✉], Charlotte R. Pennington[5], Helena Hartmann[6], Madeleine Pownall[7], Kathleen Schmidt[8], Mahmoud Elsherif[9], Nate Breznau[10], Olly Robertson[11], Tamara Kalandadze[12], Shijun Yu[9], Bradley J. Baker[13], Aoife O'Mahony[14], Jørgen Ø. -S. Olsnes[15], John J. Shaw[16], Biljana Gjoneska[17], Yuki Yamada[18], Jan P. Röer[19], Jennifer Murphy[20], Shilaan Alzahawi[21], Sandra Grinschgl[22], Catia M. Oliveira[23], Tobias Wingen[24], Siu Kit Yeung[25], Meng Liu[26], Laura M. König[27], Nihan Albayrak-Aydemir[28,29], Oscar Lecuona[30,31], Leticia Micheli[32] & Thomas Evans[33,34]

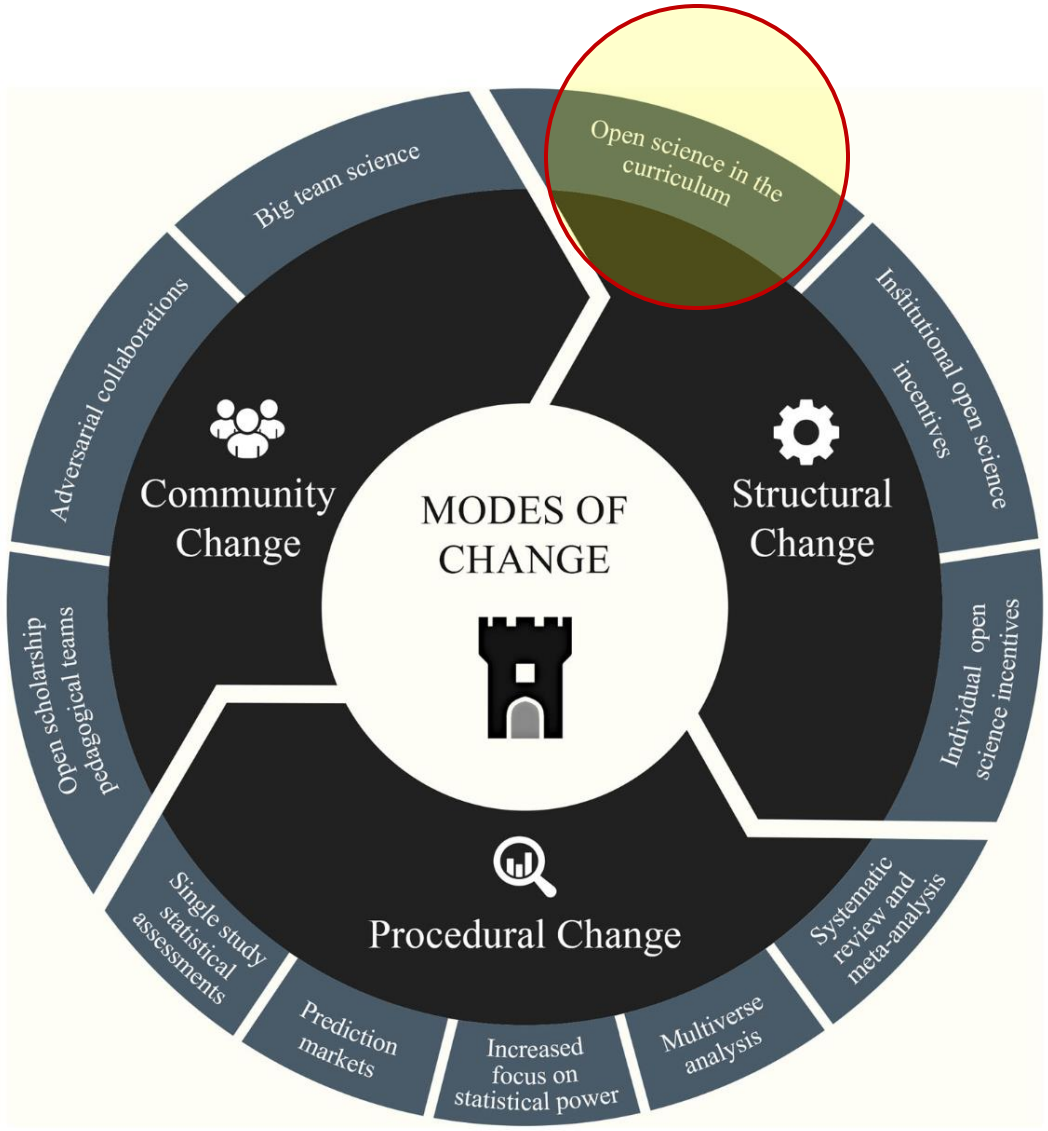

Fig. 1 Modes of change towards scientific credibility. This figure presents an overview of the three modes of change proposed in this article: structural change is often evoked at the institutional level and expressed by new norms and rules; procedural change refers to behaviours and sets of commonly used practices in the research process; community change encompasses how work and collaboration within the scientific community evolves.

II. Approaches to replication-based learning

Better practices

- Embedding open and reproducible science into teaching.
- Replication studies for Masters and PhD theses.
- Participation in the production of high-quality direct replications.

Publishing Research With Undergraduate Students via Replication Work: The Collaborative Replications and Education Project

Jordan R. Wagge[1], Mark J. Brandt[2], Ljiljana B. Lazarevic[3], Nicole Legate[4], Cody Christopherson[5], Brady Wiggins[6] and Jon E. Grahe[7]*

[1] School of Psychology, Avila University, Kansas City, MO, United States, [2] Department of Social Psychology, Tilburg University, Tilburg, Netherlands, [3] Institute of Psychology, University of Belgrade, Belgrade, Serbia, [4] Department of Psychology, Illinois Institute of Technology, Chicago, IL, United States, [5] Department of Psychology, Southern Oregon University, Ashland, OR, United States, [6] Department of Psychology, Brigham Young University-Idaho, Rexburg, ID, United States, [7] Department of Psychology, Pacific Lutheran University, Tacoma, WA, United States

Keywords: replication, pedagogy, psychology, publishing, undergraduates, teaching, projects, open science

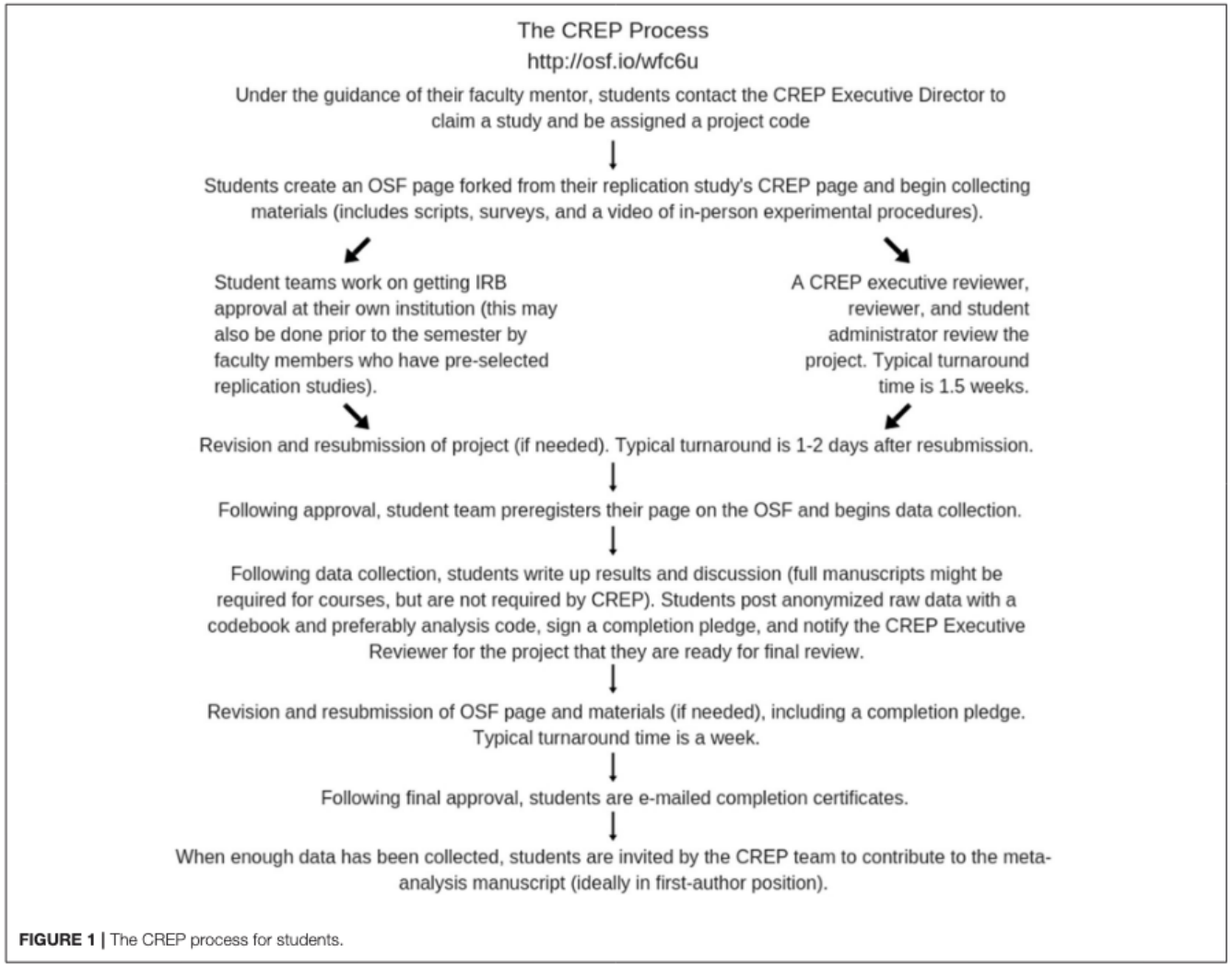

FIGURE 1 | The CREP process for students.

ORIGINAL RESEARCH REPORT

A Demonstration of the Collaborative Replication and Education Project: Replication Attempts of the Red-Romance Effect

Jordan R. Wagge[*], Cristina Baciu[†], Kasia Banas[‡], Joel T. Nadler[§], Sascha Schwarz[‖], Yanna Weisberg[¶], Hans IJzerman[**], Nicole Legate[†] and Jon Grahe[††]

The present article reports the results of a meta-analysis of nine student replication projects of Elliot et al.'s (2010) findings from Experiment 3, that women were more attracted to photographs of men with red borders (total n = 640). The eight student projects were part of the Collaborative Replication and Education Project (CREP; https://osf.io/wfc6u/), a research crowdsourcing project for undergraduate students. All replications were reviewed by experts to ensure high quality data, and were pre-registered prior to data collection. Results of this meta-analysis showed no effect of red on attractiveness ratings for either perceived attractiveness (mean ratings difference = –0.07, 95% CI [–0.31, 0.16]) or sexual attractiveness (mean ratings difference = –0.06, 95% CI [–0.36, 0.24]); this null result held with and without Elliot et al.'s (2010) data included in analyses. Exploratory analyses examining whether being in a relationship moderated the effect of color on attractiveness ratings also produced null results.

III. Our own experiences

Mario Sánchez

Ana Medrano

IPN

CICATA Legaria

Cinvestav

Master's thesis

Research in Mathematics Education, 2013
Vol. 15, No. 2, 154–164, http://dx.doi.org/10.1080/14794802.2013.797747

Routledge
Taylor & Francis Group

Are secondary school students still hampered by the natural number bias? A reaction time study on fraction comparison tasks

Jo Van Hoof*, Tristan Lijnen, Lieven Verschaffel and Wim Van Dooren

Centre for Instructional Psychology and Technology, University of Leuven, Belgium

Rational numbers and particularly fractions are difficult for students. It is often claimed that the 'natural number bias' underlies erroneous reasoning about rational numbers. This cross-sectional study investigated the natural number bias in first and fifth year secondary school students. Relying on dual process theory assumptions that differentiate between intuitive and analytic processes, we measured accuracies and reaction times on fraction comparison tasks. Half of the items were congruent (i.e., natural number knowledge leads to correct answers), the other half were incongruent (i.e., natural number knowledge leads to incorrect answers). Against expectations, students hardly made errors on incongruent items. Longer reaction times on correctly solved incongruent than on correctly solved congruent items indicated that students were indeed hampered by their prior knowledge about natural numbers, but could suppress their intuitive answers.

Keywords: dual process theory; reaction time study; natural number bias; fractions

The Natural Number Bias Hampers Secondary School Students in Different Ways: A Conceptual Replication of Van Hoof et al. (2013)

Javier Hernández Hernández | ORCID: 0009-0006-0432-6261
Instituto Politécnico Nacional, CICATA Legaria, Mexico City, Mexico
javohh@gmail.com

Ulises Xolocotzin | ORCID: 0000-0002-9896-2093
Mathematics Education Department, Research and Advanced Studies Centre, Avenida Instituto Politécnico Nacional #2508 Col. San Pedro Zacatenco, Mexico City, Mexico
Corresponding author, e-mail: Ulises.Xolocotzin@Cinvestav.mx

Mario Sánchez Aguilar | ORCID: 0000-0002-1391-9388
Instituto Politécnico Nacional, CICATA Legaria, Mexico City, Mexico

Received 26 May 2023 | Accepted 21 September 2023 |
Published online

Collaborative project

Journal of Experimental Child Psychology 113 (2012) 166–176

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Brief Report

Substitution and sameness: Two components of a relational conception of the equals sign

Ian Jones [a,*], Matthew Inglis [a], Camilla Gilmore [a], Margaret Dowens [b]

[a] *Mathematics Education Centre, Loughborough University, Loughborough LE11 3TU, UK*
[b] *Centre for Research in Applied Linguistics, University of Nottingham Ningbo, Ningbo 315100, China*

Selection criteria

- Clear theoretical background and hypotheses

Research questions

The main research question was whether the natural number bias has an intuitive character. We answered this question through the analysis of students' reaction times when confronted with fraction comparison tasks, assuming that when students compare two fractions, the intuitive answer (based on their natural number knowledge) would come to mind first. We predicted that incongruent items would trigger more mistakes (because the incorrectly intuitive answer was not inhibited through the analytical reasoning system) than congruent items (Prediction 1a). Secondly, we predicted that because inhibiting an intuitive answer requires time, there would be longer reaction times on correctly solved incongruent items than on correctly solved congruent items (Prediction 1b).

TABLE 5 Comparison of the predictions supported, not supported, and partially supported

Prediction		Original study	Replication
1a.	Incongruent items would trigger more mistakes than congruent items.	N	S
1b.	Students would take more time to answer incongruent items correctly	S	S
2.	The quantity of incorrectly solved incongruent items would decrease with age, but there would still be higher reaction times for correctly solved incongruent items in comparison with correctly solved congruent items	P	P
3.	Students would identify a conflict when they made mistakes on incongruent items, but they would not always be able to suppress the natural-number based answer	N	N

S, supported; N, not supported; P, partially supported

- Transparent report of methods and materials

Table 5
All 12 items used in the CES (conceptions of the equals sign) instrument along with group means.

Definition = means ...	Predicted conception	Group means	
		English	Chinese
... the answer to the problem	Operational	2.52	1.72
... work out the result	Operational	2.24	1.64
... the total	Operational	2.59	1.73
... that two amounts are the same	Sameness	1.64	1.86
... that both sides have the same value	Sameness	1.71	1.94
... that something is equal to another thing	Sameness	2.07	1.97
... that one side can replace the other	Substitutive	1.46	2.26
... that the right side can be swapped for the left side	Substitutive	1.47	2.13
... that two sides can be exchanged	Substitutive	1.69	2.22
... the end of the problem	Distracter	2.19	1.95
... the start of the problem	Distracter	1.35	2.24
... to repeat the numbers	Distracter	1.29	1.71

- Interest
- Impact
- Feasibility

Students' perspective

- Small survey

1. What have you learned from the replication process?
2. What advice would you give to future students who are starting a replication process?
3. Has the replication process been helpful in your studies?
4. How would you improve the replication process?

1. What did you learn from the replication process?

i. El proceso de investigación, la gestión de tiempos y la importancia de los documentos éticos

ii. Que los resultados son muy similares aún cuando la muestra es completamente diferente.

2. What advice would you give to future students starting a replication process?

i. Estudiar muy bien el documento a replicar, comprender los objetivos y laaaas metas que se plantearon, así como seguir fielmente el proceso realizado, la codificación y el análisis. Y también pensar y gestionar las escuela donde se planea aplicar los instrumentos de réplica, así como tener en cuenta los tiempos y momentos destinados a ello.

ii. Realizar el proceso con la mayor honestidad y objetividad posible.

3. Has the replication process been helpful in your studies?

i. Muy útil, es un primer acercamiento a la investigación y poder seguir paso a paso lo ya realizado por alguien más te permite experimentar vívidamente la labor de investigador siguiendo un base definida y probada.

ii. útil porque fue un primer acercamiento a una parte importante de una investigación, la toma de datos, que es algo que haré para mi proyecto de tesis.

4. How would you improve the replication process?

i. De manera personal, me hubiera gustado tener mayor precisión en la teoría (mayor explicitación en la elaboración del instrumento, no quedarnos en la sola traducción) y, una retroalimentación más cercana, para poder llegar a visualizar que los resultados obtenidos son iguales, perecidos, semejantes o diferentes, si es el caso, a los obtenidos por los autores

ii. Quizá con un banco de escuelas que estén dispuestas a colaborar en la aplicación de dichos estudios. Es muy

iii. complejo que las escuelas autoricen o tengan la voluntad de colaborar

IV. Sketching replication-based learning in Mathematics Education

Benefits

Postgraduate studies can be transformative experiences that shape students' identities

1. Learning to design and execute empirical research
2. Analyzing and interpreting replication results
3. Gaining deeper theoretical insights through constructive replications
4. Discovering how replications improve scientific progress
5. Learning to collaborate

Agenda

- How to theorize about replication-based learning in Mathematics Education?
  - Communities of practice by Lave and Wenger
- What do we want to achieve with replication-based learning in Mathematics Education?
  - Scientific identity
- What to replicate?
  - What is worth replicating and how?

Quick exploration

- Think of a study whose replication might be helpful for graduate students
- What is the key result that is worth replicating?
- What's the pedagogical value of the study?
- How would you go about it?

Take-home messages

- Adopting replication-based learning implies re-thinking our formative practices
- We need to discuss how to theorize replication-based learning
- There is a need to use our pedagogical lenses to embed replications in our curriculum

Thanks!

Ulises.Xolocotzin@Cinvestav.mx