Research in Mathematics Education
Vol. 15, No. 2, July 2013, xx-xx

GUEST EDITORIAL

Experimental methods in mathematics education research

This special issue of *Research in Mathematics Education* (RME) is devoted to research designs that involve experimental methods. Few recent studies in mathematics education have taken advantage of such methods. In fact, during 2012 only 3% of papers published in leading mathematics education journals reported experimental studies[1], a remarkably low figure. In this special issue, we aim to address this imbalance, demonstrating that careful and inventive experimental designs can illuminate a variety of questions of interest to the mathematics education community.

Specifically, experimental research designs can be of great value in our field because many of the questions that mathematics education researchers seek to address are, by their nature, causal. Interestingly, the semiotician Charles Sanders Pierce, who has had a heavy influence in mathematics education, was one of the first to pioneer experimental methods in social science research. Peirce and Jastrow's (1885) great insight was that, by using randomisation, researchers could avoid the danger of incorrectly inferring a causal relationship between two unconnected factors. Experimentation is still preferred by many empirical researchers from a range of social science disciplines for exactly this reason: it permits causal inference in a way that other research designs do not.

The structure of a basic between-subjects experiment is simple. The researcher wishes to know whether there is a causal relationship between an independent variable and a dependent variable. He or she sets up two conditions, in each of which the independent variable is set to a different level. Participants are randomly assigned to one condition or the other, and the dependent variable is measured in both conditions. Random assignment means that the two groups are probabilistically identical on all variables except one: the level of the independent variable. So, provided that the experiment is well controlled, and provided that the probability of the researcher's results occurring by chance is sufficiently low, we can be sure that it is the between-conditions difference in the independent variable that caused any observed difference in the dependent variable.

This straightforward design can reveal effects of relevance for policy-making, and part of its value in this context is in its capacity to reveal *undesirable* effects. Such was the case in a recent experiment we conducted in our own research, designed to test an educational intervention. Based on insights from the mathematics education literature, we designed multimedia presentations to combat undergraduates' well-known difficulties with understanding mathematical proofs (for details, see e.g., Alcock and Wilkinson 2011; Alcock and Inglis 2010). These presentations, which became known as e-Proofs, appeared highly successful: student feedback was remarkably positive, both in large-scale surveys and focus group settings, and lecturing staff were keen to incorporate the technology in their own classes. To quantify the effectiveness of the approach, we ran an experiment. A group of students was randomly allocated to study either an e-Proof, or to read exactly the same proof without any multimedia annotations. The results of post-test and delayed post-test

ISSN 1479-4802 print/ISSN 1574-0178 online
© 2013 British Society for Research into Learning Mathematics
DOI: 10.1080/xxxxx
http://www.tandfonline.com

measures of students' proof comprehension were clear: those in the e-Proof group performed no better than the control group at post-test, and by the delayed post-test showed significantly *worse* retention (Roy, Alcock, and Inglis 2010). What appeared to us, to other lecturing staff, and to students, to be a highly successful educational intervention was in fact detrimental to our students' learning. It took an experiment to reveal this surprising finding.

The simple logic of the experiment – that randomisation controls for all possible variables except that in which the researcher is interested (and any correlated with it) – brings a kind of aesthetic pleasure for us and for many researchers. Evans described the experimental method as "a thing of great beauty. I think of it as a tool, like a surgeon's scalpel, that can be used to dissect a psychological process and lay it bare" (2005, 23). Early in his career, he had been greatly influenced by Neisser's (1967) work on vision – not for its content, but for its compelling experimental designs:

> What fascinated me was the way in which the hypotheses were laid out and then tested so clearly and decisively in the experiments. I found the results of these experiments so convincing. They brooked of no alternative explanations (Evans 2005, 24).

Issues of design are at the forefront of this special issue, which demonstrates that designs ranging from the simple to the more complex and ingenious can be used to disaggregate effects due to cognitive, social and pedagogical factors.

Weber, for instance, used a between-subjects design in an internet study with 97 participants to demonstrate that professional mathematicians respond differently to empirical arguments according to their mathematical content. Torgerson, Wiggins, Torgerson, Ainsworth and Hewitt also used a between-subjects design, but in a study that involved 522 low-attaining 6–7 year-old children across 44 schools, and demonstrated that an intensive one-to-one numeracy programme led to improved mathematical performance on standard tests. Hegedus, Tapper, Dalton and Sloane used cluster-randomised trials with hierarchical linear modelling to compare 14–15 year-old students' learning of algebraic concepts in an innovative, technology-rich environment, with the learning in standard classroom environments.

Robert and LeFevre used a within-subjects design with 39 participants in a dual-task experiment to investigate the roles of visual and verbal working memory in young adults working on simple arithmetic. Van Hoof, Lijnen, Verschaffel and Van Dooren used a within-subjects design with 129 participants to demonstrate, via accuracy and reaction time measures, that secondary school students are hampered by a natural number bias when comparing fractions. Torbeyns and Verschaffel used a within-subjects design with 21 participants and choice/no-choice conditions to investigate strategy choice and flexibility among fourth graders working on multi-digit addition and subtraction.

This collection of empirical papers provides a picture of the way in which experimental designs can be used to address questions of importance across grade levels and across classroom- and lab-based research contexts. To extend this picture to further substantive content, we also include a review paper on the compelling issues of mathematics anxiety and stereotype threat in mathematics: Maloney, Schaeffer and Beilock review a series of studies spanning the social and cognitive psychology literatures, demonstrating that experimental methods can be useful for dissecting not only cognitive and pedagogical but also social processes. To extend the picture to further research designs, we include a book review on small-*N* experimental designs: Ridgway reviews Dugard, File and Todman's book, explaining how experimental designs can be successfully conducted and analysed with only a single participant.

We hope that this special issue as a whole demonstrates that experimental methods constitute a viable and valuable way of addressing questions of interest to mathematics educators. We believe that our field should value methodological diversity, and hope that this issue will inspire readers to consider using experiments in their own research and to submit the resultant reports to RME. We also wish to thank our authors, and especially to welcome those whose work has more often been published outside mathematics education; we believe that there is much to be learned from cognate disciplines, and we hope that this special issue will lead to greater communication and collaboration across our respective fields.

Notes

1. We surveyed the 263 articles published in 2012 by eight leading mathematics education journals (Educational Studies in Mathematics, For the Learning of Mathematics, Journal of Mathematical Behavior, Journal of Mathematics Teacher Education, Journal for Research in Mathematics Education, Mathematical Thinking and Learning, Research in Mathematics Education and ZDM). Studies were classified as being experimental if random assignment was used at the same level as the unit of analysis. A total of 8 articles fell into this category.

References

Alcock, L., and M. Inglis. 2010. Visual considerations in the presentation of mathematical proofs. *Seminar.net – International Journal of Media, Technology and Lifelong Learning* 6: 43–59.

Alcock, L., and H. Wilkinson. 2011. e-Proofs: Design of a resource to support proof comprehension in mathematics. *Educational Designer* 4. http://www.educationaldesigner.org/ed/volume1/issue4/article14/

Dugard, P., P. File, and J. Todman. 2011. *Single-case and small-n experimental designs: A practical guide to randomization tests*. 2nd ed. London: Routledge.

Evans, J. St. B.T. 2005. *How to do research: A psychologist's guide*. Hove: Psychology Press.

Peirce, C.S., and J. Jastrow. 1885. On small differences in sensation. *Memoirs of the National Academy of Sciences* 3: 75–83.

Roy, S., L. Alcock, and M. Inglis. 2010. Undergraduates' proof comprehension: A comparative study of three forms of proof comprehension. In *Proceedings of the 13th Conference on Research in Undergraduate Mathematics Education*. Raleigh, NC: SIGMAA on RUME. http://sigmaa.maa.org/rume/crume2010/Abstracts2010.htm

Lara Alcock, Camilla Gilmore and Matthew Inglis
Guest Editors