
Article:
Pampaka, M, Williams, J and Homer, MS orcid.org/0000-0002-1161-5938 (2016) Is the Educational 'What Works' Agenda Working? Critical Methodological Developments (pt 2). International Journal of Research and Method in Education, 39 (4). pp. 345-348. ISSN 1743-727X

https://doi.org/10.1080/1743727X.2016.1205817

SPECIAL ISSUE (PART 2): Is the Educational 'What Works' Agenda Working? Critical Methodological Developments

Maria Pampaka, Julian Williams, University of Manchester, Matt Homer, University of Leeds

Introduction

In this second part of this double special issue, we continue to investigate the 'What works' agenda, challenging methods and developing a debate around how best to research, and revealing implications for policy and practice. In the first part (Pampaka, Williams, & Homer, 2016) we focussed on what might be considered relatively 'orthodox' 'what works' approaches, with the dominant theme being impact evaluations and educational RCTs, along with systematic reviews and meta-analysis. Another theme was that of effective communication and dissemination in order to reach maximum impact with the relevant stakeholders.

A key aspect of our endeavour is the need on all sides of the debate to ensure that research is both rigorous and relevant, and we began asking in the first issue: Can this be achieved with current methods? In other words, can we have sufficient rigour in our methods whilst simultaneously also providing findings relevant to current educational concerns and decision-making? In this second part we develop these argument further across five new papers.

The papers in this Second issue

The first paper in this issue, authored by Kourea and Lo (2016) continues the discussion initiated with our previous issue around experimental designs and the "gold standard" of RCTs. The authors focus on another type of experimental design, the single-case design, which they claim is currently undervalued. In introducing their paper they start with some limitations of RCTs when used as "the sole criterion for determining Evidence Based practices in education" (p. 2 citing Cutspec, 2004). They point to the inability of practitioners to comply with randomization principles in everyday school conditions (in contrast to lab-based research) and suggest that the insistence on RCTs could underestimate the validity and utility of other experimental approaches not based on randomisation. An important point here is that a dogmatic insistence on RCTs as the 'one and only' research approach, i.e. one that positively suppresses other methodological forms, may – even if unintentionally - marginalise the local efforts of practitioners to improve their practice systematically taking account of their local conditions. Their paper then focuses on the presentation of the Single-Case Design (SCD) research covering its philosophical assumptions and features including a focus on the significance of the local conditions but also on robust measurement. They then review the use of SCD within the recent efforts of several leading professional organisations in social sciences and education across the U.S. and Europe. The examples they provide are drawn from the Council for Exceptional Children, the What Works Clearinghouse and the European Platform for Investing in Children: though they note the lack of reference to such designs in the last of these. After summarising the strengths and limitations of SCDs they provide some implications for policy and practice: their conclusions point to the fact that SCD

is a valuable, but under-used approach in the evidence based practice movement. For example, in efforts to 'find effective strategies for at risk-populations, it is important to focus on the behaviour of individuals' (p. 14).

The paper by Petridou and Karagiorgi (2016) (Petridou & Karagiorgi, 2016) titled "Cross-sectional predictors of 'risk' for school failure" investigates the effectiveness of a longitudinal national Programme for Functional Literacy which aims to identify students 'at risk' in mathematics and language, in primary schools in Cyprus. Its content and methodological approach offers a model case study of how research can inform and support policy and practice. They use cross-sectional multilevel analyses to identify student-level and school-level factors associated with risk of failure in language and mathematics. They also report some relatively robust psychometric approaches using the Rasch Model to validate their attitudinal and competence scales. In relation to substantive findings, they report that gender, students' confidence, sense of belonging in school and general views regarding parental involvement are associated with risk of failure for both language and maths. The paper concludes with a discussion of the implications for research and policy, focusing on the need to further explore other factors not included in the models of this study.

A discussion about potential missing evidence as well as missing data is also included in the next paper by Headley, Swoboda and Foote (2016). Their title is telling of the paper's focus and direction: "What's missing in longitudinal studies conducted in the U.S. with implications for mathematics education?". The authors delve into the topic of 'missing evidence' in longitudinal studies, highlighting the omitted evidence "related to significant research questions requiring other methods" (p. 3) as well as missing data and attrition inherent in such designs. Their motivation stems from the recognition that longitudinal studies are now "institutionalised as the best evidence for what works in mathematics education in the U.S." (p. 2). They then go on to review recent papers using longitudinal data in mathematics education in detail to provide insights into the two foci. They argue that missing data probably poses a greater threat than missing research, and they conclude with some recommendations for minimising the risk of missing evidence. Their heuristic protocol (Figure 2, p. 15) for "determining appropriate analytical techniques based on the nature of the unplanned missingness in the dataset" (p.14) should be of interest to and could benefit international educational researchers in many disciplines.

Missing data and in particular imputation techniques is also the focus of the next paper authored by Golino & Gomes (2016) titled "Random forest as an imputation method for education and psychology research: its impact on item fit and difficulty of the Rasch model". This somewhat technical paper introduces to educational research 'random forests', a non-parametric imputation technique, borrowed from the burgeoning -machine learning field. Nonetheless, there are some potentially useful pointers for all readers: for example the authors summarise very comprehensively the pros and cons of various methods for handling missing data in Table 1 (p. 2). They then go on to also to explain the random forest method through an example with the Rasch model, and conclude with a discussion of the advantages of such methods related to machine learning models and how these can applied for solving missing data problems in educational research.

The final paper in this second part of our special issue on 'what works' is authored by Kennedy, Quinn and Taylor (2016) and deals with the measurement of attitudes from a more classical test theory psychometric approach. The paper is titled "The school science attitude survey: a new instrument for measuring attitudes towards school science" and starts with a reflection on the many attempts to measure students' attitudes towards school science in the last 50 years. As the authors state, "an understanding of attitudes towards science, and how these change over time, is of particular importance to educators … who are forming a response to the continued declines in enrolments seen in post-compulsory science courses" (p. 1). The paper aims to address a gap in existing instruments providing a tool that can measure multiple facets of attitudes using a minimal number of items, while being suited for use in pre-/post-test and longitudinal studies. Those involved in longitudinal surveys where time is of the essence will definitely recognise the need for such a tool. Their paper presents a detailed validation procedure and how they applied it in a digital survey tool that led to the formation of their Science Attitude Profile. A particular strength of the resulting data analyses according to the authors is the ability to provide a comparative 'does it work?' critique of various interventions.

The issue concludes with a book review by Leguna, of the recent publication by Borgatti, Everett, and Johnson (2013) on "Analysing Social Networks". As reflected in the review, the book and the analysis presented can be a useful approach in educational research and the study of the complex interconnections between the involved 'entities' and groups.

Getting it together

Various themes and methods have been applied in the papers of this second part of the 'What works' debate. Measurement-related the Rasch model has been used as a measurement approach for measuring various aspects of a literacy programme (Petridou & Karagiorgi, 2016) and as the model for an example imputation application (Golino & Gomes, 2016). Robustness in assessing the validity of attitudinal measures has also been emphasised from a classical test theory perspective (Kennedy et al., 2016). Other advanced quantitative methods, including Hierarchical models, Longitudinal and cross-sectional data and analysis (Headley et al., 2016), Missing data and techniques to alleviate the problem (Golino & Gomes, 2016; Headley et al., 2016) have also been explored with examples from locations across the world (e.g. U.S., Brazil, Cyprus, Australia).

An important and common theme for educational research across many of these papers has been the need to continue to develop our methodology in the light of practical policy and practitioner concerns. Thus, the validity of an RCT-study that shows how an intervention programme 'works' on average across many hundreds of schools may not help a particular school, and will always need to be considered locally in the light of particular local conditions (Kourea & Lo, 2016). The argument for practitioners in the locale to continue to research and develop the specified intervention or programme will likely be very strong, and a local RCT may not be the answer, even if one agrees that robust evidence, even robust measurement, is a sine qua non.

Several of the current papers address concerns with missing data, and missing research more generally (e.g. studies of imputation algorithms, but also considerations of missing 'models', and even 'missing methods'). The recent concern of these papers with missing data expands on other work recently including some by ourselves (e.g. Pampaka, Hutcheson, & Williams, 2016) in this journal. It is becoming clear (as per Headley, et al. 2016) that many studies have failed to adequately deal with missing data, and that with the best available imputation techniques (e.g. Golino and Gomes, 2016) (though these may still be in their infancy in education) many research 'findings' would perhaps be affected. We can of course be concerned that bias in missing data may have led studies to declare programs more successful than the data would truly support if the analysis handled the missingness correctly. In general, those who 'go missing' are more likely to be those with least favourable results). But adjustments for missing data can work both ways.

Yet these considerations of course may even be amplified in the case of 'missing research': a potential problem we would like to see discussed more widely. The paper by Petridou and Karagiorgi (2016), quite typically of many papers in their acknowledgement of limitations, signals some important variables and models they would imagine further research should investigate. Thus we infer, their recommendations to practitioners and policy-makers need care: yes indeed, and how often do we take such care in trumpeting the findings of our research? When was the last time one saw the long list of important factors that were not investigated but which one might think much more important then those that were studied (often because the data was 'there', or at least easily accessible)? The almost universal use of gender as a variable may be a case in point: almost every data set has it, and it often enters models without much theoretical justification. One wonders how many research findings would be sustained if variables more theoretically justified as relevant were available to research. How many study findings would replicate (see Makel & Plucker, 2014) ?

To conclude, the really worrisome nature of 'missing research' comes from all those studies not carried out, either because educational research is so poorly funded or because practitioners largely do not see their work as research, and of course in a parallel fashion do not see 'research' as informing their day to day work. The well known publication bias (e.g. Torgerson, 2006) is, in this view, a perhaps a minor irritation compared to the bias in the research that does not get funded because it does not meet certain political, methodological, or other interests and requirements. Surely we should at least be asking: what research is missing, and why?

In the light of all this we argue that the search for new methodologies and approaches must not be declared over: the war over methods has not been 'won'. Indeed any assertion of a 'gold' or 'silver' standard feels so last century, or perhaps even more apt to the century before last.

References

Borgatti, S. B., Everett, M. G., & Johnson, J. C. (2013). *Analyzing Social Networks*. London: Sage.
Golino, H. F., & Gomes, C. M. A. (2016). Random forest as an imputation method for education and psychology research: its impact on item fit and difficulty of the Rasch model. *International Journal of Research & Method in Education*, 1-21. doi:10.1080/1743727X.2016.1168798
Headley, M. G., Swoboda, C. M., & Foote, L. (2016). What's missing in longitudinal studies conducted in the U.S. with implications for mathematics education? *International Journal of Research & Method in Education*, 1-18. doi:10.1080/1743727X.2016.1184245
Kennedy, J., Quinn, F., & Taylor, N. (2016). The school science attitude survey: a new instrument for measuring attitudes towards school science. *International Journal of Research & Method in Education*, 1-24. doi:10.1080/1743727X.2016.1160046
Kourea, L., & Lo, Y.-y. (2016). The educational validity and utility of single-case design research in building evidence-based practices in education. *International Journal of Research & Method in Education*, 1-16. doi:10.1080/1743727X.2016.1160278
Makel, M. C., & Plucker, J. A. (2014). Facts Are More Important Than Novelty Replication in the Education Sciences. *Educational Researcher, 43*(6), 304-316.
Pampaka, M., Hutcheson, G., & Williams, J. (2016). Handling missing data: analysis of a challenging data set using multiple imputation. *International Journal of Research & Method in Education, 39*(1), 19-37. doi:10.1080/1743727X.2014.979146
Pampaka, M., Williams, J., & Homer, M. (2016). Is the educational 'what works' agenda working? Critical methodological developments. *International Journal of Research & Method in Education, 39*(3), 231-236. doi:10.1080/1743727X.2016.1170476
Petridou, A., & Karagiorgi, Y. (2016). Cross-sectional predictors of 'risk' for school failure. *International Journal of Research & Method in Education*, 1-18. doi:10.1080/1743727X.2016.1189896
Torgerson, C. J. (2006). PUBLICATION BIAS: THE ACHILLES' HEEL OF SYSTEMATIC REVIEWS? *British Journal of Educational Studies, 54*(1), 89-102. doi:10.1111/j.1467-8527.2006.00332.x