Context and Implications Document for: Rethinking "quantitative" methods and the development of new researchers

Stephen Gorard
School of Education
Durham University

This guide accompanies the following article:
Gorard, S. Rethinking "quantitative" methods and the development of new researchers, *Review of Education*, [DOI will be added by Wiley]

Author's Introduction

In the UK, as in the US and elsewhere, there is considerable public investment in research capacity-building. It is clear that social science research, including education, could be improved. What is not so clear is how this improvement is to be achieved. This paper looks at what is currently the most popular and well-funded approach – the National Quantitative Methods Initiative. It shows that this initiative, very far from improving research quality, is propagating some key logical and mathematical errors. As importantly, these errors are creating misleading and perhaps dangerous research conclusions in practice. The paper suggests a simple and more rational alternative.

Implications for Policy and Practice

There is, in the UK, a national Quantitative Methods (QM) Initiative, intended to improve the quality of social science research. It is funded by the ESRC, the Nuffield Foundation, HEFCE, and the British Academy, and adds to a number of related initiatives such as the ESRC Researcher Development Initiative, and the ESRC National Centre for Research Methods (NCRM). What this paper shows is that this policy and all of the funding behind it is based on a serious logical error. We do need to address the quality of social science research but not by making more new researchers make the same mistakes currently in evidence in QM. Rather, we need to clear up and eliminate these errors before then encouraging new researchers to use numbers routinely in their research.

For example, one of the projects funded under the QM initiative collected data from two complete cohorts of undergraduate students taking two versions of a new quantitative methods teaching module. The final report to the ESRC (http://www.esrc.ac.uk/my-esrc/.../8df97b0e-a444-4074-bfc4-e4bf5d292e0e) stated that "Both cohorts reported a statistically significant reduction in statistical anxiety and a concurrent increase in self-confidence to complete statistical tasks ($p = 0.00$). This trend was more pronounced for students who completed the twenty-four week course ($p=0.00$)". Yet this data was not based on any form of randomisation. These researchers, and they are far from alone, have made an elementary mistake in their approach to analysis – using a technique predicated upon random selection but with data from populations. And these researchers are funded to help improve the quality of QM work in the UK.

Another project, funded by the same scheme, has produced resources on QM for teachers and students (http://wiserd.ac.uk/files/2613/7881/9558/QRDI-W1-ConfidenceIntervalsOneSampleTTests.pdf). It defines a confidence interval as "the level of confidence we have in our interval. For example, we can determine the interval in which we can be 95% confident that the population mean

falls". These researchers are also funded to help improve the quality of QM work in the UK, and yet their definition of a key concept is incorrect. As the accompanying paper shows, this probability is *not* what a CI portrays – it can only ever be interpreted as a long-run probability.

One of the bodies that has had perhaps the greatest beneficial impact on research policy in the UK (England) recently is the Educational Endowment Foundation (EEF). Yet even they are making the same key mistakes as the QM Initiative, encouraging the use of significance tests with non-randomised data, and allowing misreporting of attrition. For example, the EEF quality guidelines state in relation to attrition that "For cluster randomised trials, the number of clusters that dropped out will determine the threshold, not the individual pupils". This means that a school-level RCT with no school dropout would be reported as having no attrition even if 80% of the pupils in the treatment group did not provide final scores. This is clearly absurd.

At a conference to help train their picked evaluators, one of the speakers chosen by EEF stated that 'significance… indicates whether there is evidence of an… effect' (http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&ved=0CCMQFjAA&url=http%3A%2F%2Feducationendowmentfoundation.org.uk%2Fuploads%2Fpdf%2FSession_4_-_analysis_and_reporting.pptx&ei=0Xw2VMGyGIGV7AbCyID4DA&usg=AFQjCNFDqiLZLV5rnHSwUmMmlcPP8sDkaA&bvm=bv.76943099,d.ZGU). As with the CI example, this is the wrong way around. If 'significance' means anything it is the likelihood of finding a spurious 'effect' by chance even if no such effect exists in reality. Some people may think this sounds the same, but as the accompanying paper illustrates these are very different probabilities.

A paper, whose first author is the Director of the ESRC NCRM, reports 40% attrition from its sample, yet it then continues on to use significance tests, quoting standard errors, and p-values which cannot exist for what is now clearly not a random sample. It also takes no account of the missing data in the analysis, as is a widespread custom for those using significance tests with non-random samples (Patrick Sturgis, Ian Brunton-Smith, Jouni Kuha & Jonathan Jackson (2013): Ethnic diversity, segregation and the social cohesion of neighbourhoods in London, Ethnic and Racial Studies, DOI: 10.1080/01419870.2013.831932).

These examples are not the only or even the worst ones. They are used to illustrate that the problem lies at the pinnacle of social science research, and unless we solve it before extending the use of QM we will actually be using taxpayer and charitable money to make the situation worse. The accompanying paper talks about the 'vanishing breakthroughs' in all areas of science, public policy and social science, created by insecure research findings based on these statistical errors. This means that lives are being worsened, money wasted, and opportunities lost. This is a huge ethical as well as a practical and policy problem. But it could be solved almost at a stroke by the funders and publishers of research, and by those in charge of developing new researchers.