UNIVERSITY OF CANBERRA

Education Evidence Base Draft Report: STEM Education Research Centre Response

The STEM Education Research Centre (SERC) in the Faculty of Education, Science, Technology & Mathematics at the University of Canberra wishes to respond to key issues within the draft report of the Education Evidence Base Inquiry. This submission is prepared by:

- Robert Fitzgerald: Professor of Education and Senior Researcher at SERC. Rob has been a leader and innovator in the field of Information and Communication Technology Education for over 20 years working across the school, university and community development sectors. He provides leadership around technology enhanced active learning and flexible learning with a particular focus on mobile learning and the design of new learning environments. He leads the technology and learning design research program and is Director of the INSPIRE Centre for Innovation in Education and Training within the SERC.

- Simon Leonard, Associate Professor of Education, Researcher at SERC, and Visiting Scholar at Flinders University. Simon seeks to further the scientific understanding of learning through an engagement with the design and implementation of learning innovations in designed learning environments including universities, schools, online, workplaces, and science centres, including as 'Educator in Residence' at Questacon, the Australian National Science and Technology Centre. Simon is a foundation Fellow of the Institute for Mixed Methods Research. He has previously held roles as Head of Teacher Education, and Academic Coordinator of the Ngunnawal Indigenous Higher Education Centre, both at the University of Canberra.

Our response to the Draft Report comments on the use of randomised controlled trials (RCTs); reliance on normalised testing to measure education outcomes; and the inclusion of data on teacher uptake and implementation of research outcomes in the consideration of education evidence.

Prior to addressing specific paragraphs within the report, we wish to emphasise that we support the development of a quality education evidence base, and that this evidence base should be focused on serving the needs of students, not only researchers. We also endorse the Commission in its findings that wider access to datasets should be encouraged and enabled. Improved access to the datasets developed by researchers would ensure that evidence on what works for different students in different settings does not get lost when data is aggregated and averaged. This information is key to understanding the value add of educational programs and essential to informing successful differentiated educational designs. Given most educational research is publically funded, we support any finding that encouraged – even expected – researchers to share their datasets.

Overall, however, we emphasise the importance of avoiding creation of a repository of education evidence for which the primary purpose is a judgement around the success or failure of the national education system, based on an

T +61 (0)2 6201 5111
W www.canberra.edu.au

University of Canberra
Locked Bag 1
ACT 2601 Australia

assumption that if we measure something, it will improve. Instead, the purpose of accumulating this evidence should be to identify what works for individual students and groups of students in particular contexts, and on informing iterative cycles of education design and implementation . For this purpose, data created through non-experimental research is an important source. As Slavin (2004) emphasises,

> *for questions that compare the outcomes of alternative programs or policies, there is no substitute for a well-designed experiment.*

but

> *For many questions, non-experimental methods are perfectly appropriate.*

In its inquiry into the education evidence base, we strongly commend to the Commission the growing international body of work into design-based implementation research (Penuel, Fishman, Cheng, & Sabelli, 2011) which is exploring how – and what kind of – evidence and research improves student outcomes and leads to scalable, sustainable change in classroom practices.

Draft Report: The Commission supports investment in high-quality research, particularly randomised controlled trials, to build the Australian evidence base on what works best to improve education outcomes. (p. 168)

The challenge of both doing and interpreting educational research is that learning is an inherently complex activity that is influenced by dozens, if not hundreds of variables that act in combination rather than isolation (Musso, Kyndt, Cascallar, & Dochy, 2012). In the face of this, the ranking of effect sizes from the meta-analysis of randomised controlled studies (Hattie, 2009) can be alluring to educators. However, while ranking of effect sizes offers a direction for educational design, it does not provide the level of differentiated design detail required to understand 'what works, for whom and under what conditions'. Even in health research, however, the primacy of this research approach has been strongly challenged for some time. Feinstein and Horwitz (1997), for example, argued for a clear differentiation of evidence-based practice and 'best available' evidence. They noted that the gold standard of the RCT showed efficacy for the 'randomised' average patient, but 'not for pertinent subgroups formed by such cogent clinical features as severity of symptoms, illness, co-morbidity, and other clinical nuances'. Further they found that randomised trial information is

> *…seldom available for issues in etiology, diagnosis, and prognosis, and for clinical decisions that depend on pathophysiologic changes, psychosocial factors and support, personal preferences of patients, and strategies for giving comfort and reassurance.*

Other health research has even questioned the assumption that the RCT is actually superior at measuring the things it can measure such as a study by Concato, Shah, and Horwitz (2000) which used meta-analysis to conclude that, in contrast to the common assumption:

> *The results of well-designed observational studies (with either a cohort or a case–control design) do not systematically overestimate the magnitude of the effects of treatment as compared with those in randomized, controlled trials on the same topic.*

As is often the case in health, educational researchers are finding that the evidence that can be collected through RCTs - that is findings about relatively straightforward interventions where much is constant - is not always sufficient evidence for action. In their work using neural network analysis rather than regression methods, for example, Musso et al. (2012) have found that the successful predictive models of mathematical performance for high, medium and low performing students are

actually built on different combinations of variables. That is, they were not able to model the average student, although they were able to make reasonable models for sub-groups of students.

In seeking to collect education evidence, we therefore caution against an assumption that simple best-practice exemplars can be identified. Educational research provides useful guidance, but it cannot offer universal solutions for the design of educational programs and environments. Rather than seeking universal answers, educational evidence should be used to identify design principles that can then be adapted to context.

Draft Report: A range of school education outcomes are assessed through the National Assessment Program (NAP) … (p. 83)

We would contend that reliance on the NAP as a measure of student progress may not capture nuances of learning which are relevant to effective learning design, and that reliance on such datasets may lead to missed opportunities. NAP-style – and indeed RCT – measurements are often used to support judgements on the success or failure of the national education system. They seldom, however, prove useful in informing education design choices and implementation, or iterative evaluation and improvement of what works for students in specific contexts. We use different findings on the effect size of using problem based learning (PBL) as an example.

Hattie's (2009) meta-analysis work is widely known in Australia and its conformity with the so-called gold standard research methodology has made it highly influential. By combining the findings of the many studies included in his analysis, Hattie finds that PBL has an effect size nearly indistinguishable from the general effect of schooling, and far lower than the effect of direct instruction. Hattie's analysis is based on normalised or pre/post testing which rarely assess capacity to apply learning in novel situations, and certainly do not assess abilities in an enacted way. Research such as Hattie's essentially assumes that as long as assessment is well designed, it is neutral.

An alternative meta-analysis on PBL (Dochy, Segers, Van den Bossche, & Gijbels, (2003), however, has found that the effect size of PBL is in fact strongly affected by the nature of the test. Specifically, they found that the effect size of PBL was much larger when the learning was assessed using methods requiring greater retrieval effort such as short answer or free recall questions, when compared to assessment based on recognition tasks such as multiple choice tests. The Dochy et al. study also finds that while the amount of knowledge learned through PBL seems to be slightly less than through methods such as direct instruction, the knowledge that is learned appears to be held for longer and to be more readily accessed.

This comparison demonstrates that even when educational research investigates similar problems and employs similar methodologies, educational design based on the findings of research could look quite different. In this case Hattie's research might see an educational designer avoiding PBL, while the same designer might adopt PBL if only Dochy et al. had been read.

The tension is not in the findings, but in the goals and measures of the research. When these two studies are taken together, the design principle that emerges is that PBL is not useful if the goal is rapid gains in knowledge, but that PBL is a useful approach if the goal is longer-term knowledge retention, with the advantage of PBL appearing to lie in its support for learners to develop better organising structures for new knowledge. The choice to use PBL or not is then a design decision made relative to the goal of the designer and the learners in the program they are designing. It would be a poor choice, for example, for a bridging course where the primary goal is to achieve a specified grade in order to be admitted to a degree course. However PBL would appear a much better choice in the actual degree course where the goal is increased graduate competence.

Draft Report: Crucially, evidence can only support improved outcomes as well as transparency and accountability to the extent that it is accessed and used by relevant decision makers in the education system. Even the highest-quality evidence based on the most rigorous analysis cannot improve education outcomes if it does not find its way into classroom practices in schools and ECEC services. (p.65)

Overall, the Commission concludes that initiatives to improve the use of evidence in the design of education policy and programs are warranted. (p. 178)

The draft report focusses primarily on the education evidence which can be used to show improved education outcomes. The Terms of Reference, however, require the Commission to report and make recommendations on 'the information required to provide a comprehensive evidence base to inform policy development in early childhood and school education now and in the future'.

Research into how (and if) practitioners translate research findings and training into changed classroom practice is a relatively young field in education research. It is, however, an improvement on previous research which has focused on how satisfied practitioners are with professional development activities, as it investigates instead the links between continuing professional development (CPD) activities and changes in classroom practice (Penuel, Fishman, Yamaguchi, & Gallagher, 2007).

As an example, there are now numerous projects where researchers, museums and other institutions have helped to develop inquiry-oriented or enacted curricula designed for teachers (Davis & Krajcik, 2005) and similarly numerous programs offering teacher professional development in inquiry approaches. While many of these projects are highly successful internally, two commonly reported challenges are scaling and sustainability. Even where research or other projects show clear success, there is little uptake beyond the project by others not involved in the project, or even by teachers who are involved in the project but do not come to 'own' the intervention as part of their regular pedagogical practice (Coburn, 2003).

Given the highly situated nature of cognition and learning, these challenges are not a surprise. It has been understood for many decades now that it is teachers' adaptations and the variations in environments, rather than the plans of policy makers, curriculum writers or researchers, that determines the effectiveness of an educational program (Braun, Ball, Maguire, & Hoskins, 2011; Elmore, 1979; McLaughlin, 1987). And once the 'hothouse' of funding, resources, or simply the focus a project brings is gone, the everyday demands of school life return and innovation often falls away (Fishman, Penuel, Hegedus, & Roschelle, 2011).

Research on scaling and sustaining educational innovation is also in its infancy, but already a good deal is known. Some key principles that can be identified include that CPD must be coherent with teachers' own development goals, the goals they have for their students, and the demands of their employer (Janssen, Westbroek, & Driel, 2014; Penuel et al., 2007); that reform-like professional learning is more effective in creating sustainable changes to teacher practice (Garet, Porter, Desimone, Birman, & Yoon, 2001); and that teachers need to see that classroom innovations are congruent with the multiple goals they seek to achieve simultaneously in the classroom, including optimising learning, covering content, sustaining student willingness to participate, and promoting classroom behavioural norms (Doyle & Ponder, 1977; Janssen, Westbroek, & Doyle, 2015).

We contend that research and evidence showing how practitioners translate findings on 'what works' into classroom practice, and the research and evidence supporting scaling and sustaining educational innovation are key components of the education evidence base, and should be included

if the Commission is to meet its terms of reference with regard to providing 'a comprehensive evidence base to inform policy development in early childhood and school education now and in the future'.

References

Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011). Taking context seriously: towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education, 32*, 585-596. doi: 10.1080/01596306.2011.601555

Coburn, C. E. (2003). Rethinking Scale: Moving beyond Numbers to Deep and Lasting Change. *Educational Researcher, 32*(6), 3-12. doi: 10.3102/0013189X032006003

Concato, J., Shah, N., & Horwitz, R. I. (2000). Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs. *New England Journal of Medicine, 342*(25), 1887-1892. doi: 10.1056/NEJM200006223422507

Davis, E. A., & Krajcik, J. S. (2005). Designing Educative Curriculum Materials to Promote Teacher Learning. *Educational Researcher, 34*(3), 3-14. doi: 10.3102/0013189X034003003

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction, 13*(5), 533-568. doi: 10.1016/S0959-4752(02)00025-7

Doyle, W., & Ponder, G. (1977). The ethic of practicality and teacher decision-making. *Interchange, 8*, 1-12.

Elmore, R. F. (1979). Backward Mapping: Implementation Research and Policy Decisions. *Political Science Quarterly, 94*(4), 601-616.

Feinstein, A. R. M. D., & Horwitz, R. I. M. D. (1997). Problems in the "evidence" of "evidence-based medicine". *The American Journal of Medicine, 103*(6), 529-535. doi: 10.1016/S0002-9343(97)00244-1

Fishman, B. J., Penuel, W. R., Hegedus, S., & Roschelle, J. (2011). What happens when the research ends? Factors related to the sustainability of a technology-infused mathematics curriculum. *Journal of Computers in Mathematics and Science Teaching, 30*(4), 329-353.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What Makes Professional Development Effective? Results from a National Sample of Teachers. *American Educational Research Journal, 38*(4), 915-945. doi: 10.3102/00028312038004915

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* New York: Routledge.

Janssen, F., Westbroek, H., & Driel, J. (2014). How to make guided discovery learning practical for student teachers. *Instructional Science, 42*(1), 67-90. doi: 10.1007/s11251-013-9296-z

Janssen, F., Westbroek, H., & Doyle, W. (2015). Practicality studies: How to move from what works in principle to what works in practice. *Journal of the Learning Sciences, 24*(1), 176-186. doi: 10.1080/10508406.2014.954751

McLaughlin, M. W. (1987). Learning from Experience: Lessons from Policy Implementation. *Educational Evaluation and Policy Analysis, 9*(2), 171-178. doi: 10.2307/1163728

Musso, M., Kyndt, E., Cascallar, E., & Dochy, F. (2012). Predicting Mathematical Performance: The Effect of Cognitive Processes and Self-Regulation Factors. *Education Research International, 2012*, 1-13. doi: 10.1155/2012/250719

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What Makes Professional Development Effective? Strategies That Foster Curriculum Implementation. *American Educational Research Journal, 44*(4), 921-958. doi: 10.3102/0002831207308221

Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing Research and Development at the Intersection of Learning, Implementation, and Design. *Educational Researcher*, 40(7), 331-337. doi: 10.3102/0013189X11421826

Slavin, R. E. (2004). Education Research Can and Must Address "What Works" Questions. *Educational Researcher*, 33(1), 27-28. doi:10.3102/0013189x033001027