

Wools, S., Eggen, T., & Sanders, P. (2010). Evaluation of validity and validation by means of the argument-based approach. *Cadmo*, **18**, 1, 63–82.

Wright, B. & Linacre, J. (1994). Reasonable Mean-square Fit Values. *Rasch Measurement Transactions*, **8**, 3, 370.

Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis*. Chicago: Mesa Press.

Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In: C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. Psychometrics*. 45–79. Amsterdam, Netherlands: Elsevier Science.

Afterword

It has been eight years since the publication of this special issue exemplifying 'An approach to validation' (and closer to ten years since the work it describes was conducted). Validation studies continue to be demanding activities, not helped by considerable variety in views about what validation should involve, what it can achieve and whom it should serve (Newton & Shaw, 2016). One thing is clear, however. There is an increasing demand for awarding bodies to demonstrate the quality of their qualifications and meeting this demand is no mean feat.

Our main motivation for publishing this work was to provide a practical example for would-be validators by describing the framework (based on Kane, 2006) and methods that we applied in a validation study of International A level Physics. More detailed description of some elements of the study can be found in Crisp and Shaw (2012). An earlier pilot validation study is described in Shaw and Crisp (2010a; 2010b), whilst use of the literature to develop the framework is described in Shaw, Crisp and Johnson (2012).

Since the study on A level Physics, researchers at Cambridge Assessment have conducted validation studies for a variety of other qualifications (e.g. IGCSEs and International A levels). Some elements of these studies have been reported in various publications and at conference (e.g. Grotorex & Shaw, 2012; Grotorex et al., 2013a, 2013b). Over time, we have made some adjustments to the set of methods used, jettisoning a small number of validation methods that were resource-intensive but provided minimal additional validity evidence, and adjusting or extending others – for example, to gather validity evidence about speaking assessments. Where methods have been changed, care has been taken to ensure that the revised set of methods still provides evidence in relation to each validation question.

Five years after implementing the original validation framework, a number of issues emerged which prompted us to review and revise the framework, as described in Shaw and Crisp (2015). We believe that these changes have strengthened the theoretical structure underpinning the framework. The changes were made to validation questions 4 and 5.

Validation question 4 relates to the *Extrapolation* inference and was previously phrased in terms of whether the constructs sampled are representative of competence in the wider subject domain. We broadened the question to include related competence beyond the subject. In the revised framework (Shaw and Crisp, 2015) it appears as:

Do the constructs sampled give an indication of broader competence within and beyond the subject?

The *Decision-making* inference was revised to better reflect current thinking (e.g. Kane, 2013). Appropriate decisions can only be made if the meaning of test scores is clearly interpretable by a range of

relevant, credible stakeholders. However, the previous wording of the validation question for this inference focused too much on providing guidance to stakeholders on the meaning and uses of results, and not enough on whether scores and grades indicate students' potential. Validation question 5 appears in the revised framework as:

Do scores/grades give an indication of success in further study or employment such that they can be used to make appropriate decisions?

Since the revision of the framework, a number of new methods have been explored in order to address the changes. For example, we have used the size of the correlation between student results in a specific IGCSE and later performance in AS or A levels to provide evidence relating to validation question 5 for that IGCSE.

As described in the 'Conclusions' section of the Special Issue, a key challenge with validation work is the breadth and depth of evidence needed when conducting a study of the kind described. Thus, we alluded to how it might be appropriate to implement the full validation approach to a small number of qualifications, and to apply a more streamlined, *operational* approach to validation to some further qualifications. This was discussed in Shaw and Crisp (2011) and has since been implemented for a range of qualifications. The *operational* approach uses the same validation framework but uses only existing, operationally-available data such as marking (scoring) data and documentary evidence (as opposed to data generated through experimental work). This *operational* approach may not be able to address each of the validation questions as robustly as the original *experimental* approach. However, conducting some studies of each type allows awarding bodies to provide validation evidence for a wider range of qualifications.

Following on from the development of the *operational* approach, a hybrid of the two approaches (operational and experimental) has since been trialled. This involved routinely available evidence plus gathering some new data using a small number of methods from the experimental approach. Relevant stakeholders selected those methods of particular interest to their assessment context. Whilst a hybrid approach will not be as substantive as a full experimental study, it may nevertheless yield targeted validity information in a more time and cost effective way. Balancing the robustness of evidence against the resources involved in its collection continues to be an ongoing debate in the implementation of validation studies.

Given that the quality of qualifications needs to be ensured, we would still argue that "the challenge of validation – no matter how great, should not impede its continuing execution" (Shaw & Crisp, 2015, p.36).

References

- Crisp, V., & Shaw, S.D. (2012). Applying methods to evaluate construct validity in the context of A level assessment, *Educational Studies*, **38**, 2, 209–222.
- Greatorex, J., & Shaw, S.D. (2012). The validity of teacher assessed Independent Research Reports contributing to Cambridge Pre-U Global Perspectives and Research. *Research Matters: A Cambridge Assessment Publication*, 14, 38–41.
- Greatorex, J., Shaw, S.D., Hodson, P., & Ireland, J. (2013a). *Do the questions from A and AS level Economics exam papers elicit responses that reflect the intended construct?* Poster presented at the British Educational Research Association annual conference, Brighton, 3–5 September.
- Greatorex, J., Shaw, S.D., Hodson, P., & Ireland, J. (2013b). Using scales of cognitive demand in a validation study of Cambridge International A and AS level Economics. *Research Matters: A Cambridge Assessment Publication*, 15, 29–37.
- Kane, M.T. (2006). Validation. In: R.K. Brennan (Ed.), *Educational Measurement* (4th ed.). Westport: American Council on Education/Praeger.
- Kane, M.T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, **50**, 1, 1–73.
- Newton, P.E., & Shaw, S.D. (2016). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, **23**, 2, 178–197.
- Shaw, S.D., & Crisp, V. (2010a). Developing and piloting a framework for the validation of A levels. *Research Matters: A Cambridge Assessment Publication*, 10, 44–47.
- Shaw, S.D., & Crisp, V. (2010b). *How valid are A levels? Findings from a multi-method validation study of an international A level in geography*. Paper presented at the Association for Educational Assessment in Europe, 11th annual conference, Oslo, 4–6 November.
- Shaw, S.D., & Crisp, V. (2011). *Translating validation research into everyday practice: issues facing an international awarding body*. Paper presented at the Association for Educational Assessment in Europe, 12th annual conference, Belfast, 10–12 November.
- Shaw, S.D., & Crisp, V. (2015). Reflections on a framework for validation – 5 years on. *Research Matters: A Cambridge Assessment Publication*, 19, 31–37.
- Shaw, S.D., Crisp, V., & Johnson, N. (2012). A framework for evidencing assessment validity in large-scale, high-stakes international examinations. *Assessment in Education: Principles, Policy & Practice*, **19**, 2, 159–176.