

Editorial

In place of the more usual collection of diverse takes on key issues in assessment and assessment research, this Special Issue of *Research Matters* consists of something more akin to a monograph – offering an overview of an in-depth empirical validation study. Far from being mere pragmatic empiricism, this study is driven by a complex mix of theoretical underpinning, practical validation, and ethical concern.

Although an imperfect metaphor, one can sometimes think of validity as a mysterious, invisible force acting on assessment – much as those in the seventeenth and eighteenth centuries regarded gravity: ‘...But shall gravity be therefore called an occult cause, and thrown out of philosophy, because the cause of gravity is occult and not yet discovered?’ (Roger Cotes, Preface to Isaac Newton’s *The Mathematical Principles of Natural Philosophy*, 1729).

Validity is not an obvious, readily observable quality of individual qualifications and assessments. During the Enlightenment, the intangibility of gravity, magnetic fields, electrical current and so on – things visible in their effects rather than in their form – posed a serious challenge to scientific explanation and enquiry. Likewise, validity is an elusive quality of any assessment and of the inferences made from the assessment outcomes. Intangible it may be, but unlike gravity, validity is located in complex social systems, and thus poses considerable challenges, both in respect of definition and in respect of practical validation methodology. Historical transnational analysis of validation, currently being undertaken by Paul Newton and Stuart Shaw (the latter being co-author of this Special Issue), indicates the highly contested and ‘slippery’ nature of working definitions of validity. The dispute over the extent of engagement with the uses of an assessment implied by Samuel Messick’s propositions of the 1990s marks merely a recent contestation of the limits and nature of validity and validation work. Newton and Shaw’s historical analysis suggests that fundamental schism in validity theory has been present since it became an object of interest by assessment theorists and developers – it is not a recent ‘crisis’.

So, if the Trinitarian formulation of validity is disputed, if its elaboration into ‘many validities’ appears as unmanageable post-modernism, if the Consequentialist doctrine is unstable, are there adequate grounds for undertaking ambitious validation studies of the kind outlined in this Issue? We believe that there are. The Cambridge Approach emphasises the moral obligations falling on those developing, administering and evaluating assessments and qualifications; since assessments are intended to convey personal, social and economic goods, and depend on and create trust, then the claims which are made for an assessment or qualification must be justified. This reflects Kane’s concept of ‘validity argument’; in other words, there remains a pragmatic and ethical basis for engaging in validation effort. Claims are made by those who develop and use assessments: that they attest to possession of knowledge, understanding and/or skills; that they predict later attainment or performance; that they adequately prepare a person for an occupation or for participation in an educational programme.

The Cambridge Approach – and the study reported here – reflects growing expectations by all users of specific assessments and qualifications that such claims are sufficiently warranted.

The origins of the study lie in fundamental work which members of

Cambridge Assessment have had in train for the last three decades. Cambridge ESOL (English for Speakers of Other Languages) transformed its practice by ensuring that extensive work was undertaken on the precise nature of ‘the construct’ – the ability to functionally use English, as a speaker of another language. This approach placed validity and validation at the heart of assessment development and of assessment practice. Rather than only retrospectively analyse its assessments, this reorientation of underpinning theory and of development practice focused on understanding fully, and with precision, the nature of the construct. The justification for the claims made by the qualifications became the starting point for development and operations. Validity was placed at the heart of qualifications, design, operation and evaluation. When I joined Cambridge Assessment in 2006 it seemed prudent to take this uncompromising focus on validity, developed in the context of language testing, out to general qualifications such as GCSE and A level, where more eclectic development processes tended to be used, not least as a result of regulatory demands.

International qualifications carry particular complexities. They need not only to be tuned to the values, educational forms and complex functions existing in specific national contexts, they also need to support international ‘exchange value’, for example, for the purpose of gaining access to Higher Education in a country other than the one in which the learning and assessment for the qualification took place. The ‘validity claims’ are thus complex and extensive. The study reported in this volume is a direct response to the challenges that this poses. As this study was in progress, increasing questions were being asked about the pressures being exerted on qualifications from the use of qualifications in accountability mechanisms and other ‘control’ elements of the education and training system. Even if we do not have a fully articulated ‘consensus’ definition of validity – nor simple agreement on the scope of validation studies – the authors of the study have responded to the ethical and practical issues of justifying both the claims made, and the uses to which the qualification is put.

This study was also an important test case – just what would a substantive validation study look like? How much resource would it consume? Victoria Crisp and Stuart Shaw not only used contemporary theory to scope this study, but also took extensive soundings in the research and development community regarding utility and probity. Gordon Stobart is right in describing this as a ‘Rolls-Royce study’. It is a study in the art of the possible – not restricting itself from the outset, but exploring just what a wide-ranging study would consist of. The work was exhausting as well as exhaustive.

Our transnational and domestic policy work predicts that expectations regarding validity – amongst candidates, parents, users of results, funding agencies etc – looks set to increase. This study in the ‘art of the possible’ enables us to develop protocols for well-grounded, tractable and informative validation work – fine-tuning studies to purpose, resource and opportunity.

Tim Oates

Group Director

Assessment Research and Development

The Cambridge Approach:

<https://www.cambridgeassessment.org.uk/cambridge-approach/>