WORKSHOP TITLE:

Issues around how best to provide evidence for assessment validity: the challenge of validation

Presenters:

Stuart Shaw and Sarah Hughes

Presenters' Bios (500 words):

Stuart Shaw
Principal Research Officer – Assessment Division, Cambridge International Examinations (Cambridge Assessment)

Stuart began his career as an engineer, and holds an honours degree in Physics, a diploma in Applied Physics and a research degree in Metallurgy. His early experience, gained with an international plc, covered a range of engineering specialisms. Following his time in industry, he entered the TEFL world (Teaching English as a Foreign Language), gaining a certificate and diploma in TESOL and a Master degree in Applied Linguistics. He had several years experience as an EFL teacher and Director of Studies. Stuart also holds a postgraduate degree in Theology.

Stuart has worked for Cambridge Assessment since January 2001 where he is particularly interested in demonstrating how Cambridge Assessment seeks to meet the demands of validity in its assessments. Before leading a research team in the area of mainstream international examinations, Stuart worked on a range of Cambridge English (formerly Cambridge ESOL - English for Speakers of Other Languages) products with specific skill responsibilities for assessing writing. He has experience in the areas of researching and managing second language writing assessment in an ESOL context; developing, revising and introducing assessment procedures for new testing scenarios and disseminating good practice to others through mentoring, lecturing, informal advice; and the establishment of criteria for good practice in the type of public examination offered by University of Cambridge English Examinations and Cambridge International Examinations.

Stuart is an experienced presenter and has lectured for the School of Fuel Management and for the Department of Theoretical and Applied Linguistics (University of Cambridge). He is currently an affiliated lecturer with the Faculty of Education (University of Cambridge).

Stuart has a wide range of publications in English second language assessment and educational research journals. Recent books include: Examining Writing: Research and practice in assessing second language writing (Shaw & Weir, 2007); The IELTS Writing Assessment Revision Project: towards a revised rating scale (Shaw & Falvey, 2008) and Validity in Educational and Psychological assessment (Newton & Shaw, 2014). He is currently

working on his next book: Language Rich: Insights from Multilingual Schools (Cambridge University Press).

Stuart is a Fellow of the Association for Educational Assessment in Europe (AEA-Europe) and member of the Professional Development Committee (AEA-Europe). He is also a Fellow of the Chartered Institute of Educational Assessors (CIEA).

Sarah Hughes
Sarah is a Researcher at Cambridge International Examinations. Her background includes a first degree in applied psychology and a Master's degree in educational assessment. Her career roles have involved primary school teaching, national test development and assessment-related educational research. Her journey to Cambridge International Examinations involved roles at Cambridge Assessment, The UK Qualifications and Curriculum Development Authority and Pearson Education.

Why AEA members should attend this workshop:

The workshop is intended to make the complexities of validation theory and practice more apparent and more understandable (hopefully).

Who this Workshop is for:

The workshop is envisaged as a resource for students in educational measurement and assessment, for key practitioners in assessment agencies who wish to gain a deeper understanding of validation, for those with an academic interest in assessment, and for the validity novice who should be able to benefit from attending the workshop.

Overview (500 words):

"The provision of satisfactory evidence of validity is indisputably necessary for any serious test" (Hughes, Porter & Weir, 1988, p.4)

Given the fundamental aim of ensuring that educational assessments are valid, it is important for assessment agencies to be able to provide evidence of the validity of their assessments. The commitment of assessment agencies to continually improve validity is both encouraging and increasingly more necessary in a culture of accountability.

Assessment providers are increasingly recognising the benefits of demonstrating to stakeholders that the ability constructs they are attempting to measure are well grounded in the assessments they offer. However, as Messick points out, not all testing agencies take validity seriously: "many test makers acknowledge a responsibility for providing general validity evidence of the instrumental value of the test but very few actually do it" (1992, p.89).

How can this be?

Unfortunately, the *Standards for Educational and Psychological Testing* (2014) gives no prescription or guidance regarding specific types of evidence to gather and how to gather it; no guidance on the types of evidence of validity that do not fit into the five prescribed categories of evidence; no examples or references to an 'adequate' validation study; no statistical guidance (e.g. Differential Item Functioning, statistical analysis of test bias); little clarity on validity evidence for accountability testing; and, no advice on how to evaluate validity arguments.

The *Standards* suggests that validation requires a complex web of evidence and analysis: a many-linked chain. But, does that mean that, if certain links in the argument chain are missing, then a claim of validity cannot be made? Do even the largest assessment agencies really produce that much evidence and analysis on a routine basis? And, if not, then does that mean that they cannot really claim validity? If an assessment agency failed to provide certain strands of evidence, would that prevent it from claiming that its tests are valid? When it comes to validation, how little evidence can we rely upon, should we rely upon, do we rely upon, in order to support the strength of a validity claim? What counts as evidence of validity in practice anyway? And how much, of what kind, of empirical evidence and/or logical analysis is required in order to stake a claim to validity? Quite apart from where to start, how to proceed and when to stop a validation study, there is also the thorny issue of how to report its findings and to whom. And what about the criteria for evaluating a validation argument? How are they to be understood and applied?

So many questions.

This workshop will highlight the challenges faced when validating the intended interpretation of test scores and their relevance to the proposed uses of those scores. It is hoped that the workshop will engender discussion that will focus on the issues raised when developing, piloting and implementing (in an operational context) a test validation framework which attempts to structure validity evaluation via a number of questions representing components of validity for specific qualifications.

Validating proposed interpretations and uses of test scores is a difficult task and it is hoped that by sharing experiences through a collaborative workshop environment, greater insights will be drawn leading to an increased understanding of the validation process.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Hughes, A., D. Porter, and C.J. Weir. 1988. Validating the ELTS test: a critical review. Cambridge: The British Council and the University of Cambridge Local Examination Syndicate.

Messick, S. 1992. Validity of test interpretation and use. In Encyclopedia of educational research (6th ed.), ed. M.C. Alkin. New York: Macmillan.

Preparation for the workshop:

Useful preparatory reading:

Shaw, S. & Crisp, V. (2012). An approach to validation: Developing and applying an approach for the validation of general qualifications. *Research Matters*, Special Issue 3, 1-44. (a copy will be provided for workshop participants)

Chapter 5 - The deconstruction of validity: 2000-2012 in Newton, P. E & Shaw, S. D. (2014). Validity in Educational and Psychological Assessment. London: SAGE

Schedule

The schedule given here is tentative but will broadly follow the following:

Time	Session	Presenter
09.00	Coffee and registration	Yet to be decided
09.30	Introductions: why is validation so important?	
09.45	Theoretical perspectives on validation Developing a validation framework for general educational assessments	
11.00		
11.30 (with 30 min break for lunch)	Validation challenges • Claims, purposes, interpretations, uses • Constructing validation arguments • Identifying and collecting validation evidence (adequacy, and relevance) • Evaluating validation arguments • Reporting validation findings	
15.30		
15.45	Feedback on issues discussed in relation to particular contexts	
16.30	Summary of issues	