

Making the case for a good assessment

Darren Conway



Validation means building a case

*“...broadly speaking, validity is nothing less than an evaluative summary of both the evidence for and the actual as well as potential consequences of score interpretation and use ... This comprehensive view of validity integrates considerations of content, criteria, and consequences into a construct framework for empirically testing rational hypotheses about score meaning and utility. Fundamentally, then, score evaluation is empirical evaluation of the meaning and consequences of measurement. As such, **validation combines scientific inquiry with rational argument to justify (or nullify) score interpretation and use.**” (Messick 1994: 3)*

The most relevant English NZ standards

- 2.2 The school's assessment processes and decisions are open and consistent, and its assessment methods are fair, valid and appropriate
- 2.2m For high-stakes assessments such as those bearing on entry to mainstream education programmes, • administration of the assessment is appropriately secure; and • moderation assures consistency of assessments; and • the assessment is benchmarked against other acceptable assessments; and • there is robust evidence for any claims of equivalence.

Valid

- The assessment can be interpreted as representing ability to cope with the communication demands of English-medium university study at the appropriate level
- The assessment reflects course content
- Assessment tasks, including the conditions under which they are performed, cover a relevant and sufficient sample of those typical at university
- Texts and language represent academic use
- Judgement criteria and scoring accurately reflect the relevant aspects of ability defined in the construct

Relevant evidence might include

- A construct definition
- A reference description/analysis of Target Language Use
- A course outline
- Text analysis (vocabulary, lexical density, etc)
- Expert judgement of appropriacy of texts, tasks & rating (e.g. pre-moderation)
- Clear and firm re-submission policy

Benchmarked

Initial application, at least:

- At content level against established language proficiency frameworks such as CEFR, Pearson GSE
- At assessment task level, of rubrics, grading scales, processes

Within 12 months, the above plus:

- At grading level so that decisions are comparable with other accepted measures (e.g. IELTS, TOEFL, PTE, NZCEL)
- Of outcomes for successful assessment candidates once they begin university study

Relevant evidence might include

- Content and process comparison
- Benchmarking policy and procedures
- Parallel score evidence
- Tracer studies e.g. from Alumni, including follow-up student and/or lecturer questionnaires, GPA, pass rates, etc.

Reliable/consistent

Initial application, at least:

- Assessment scores are consistent and comparable across different administrations of the assessment and across different groups of students and different raters.

Within 12 months, the above plus:

- Assessment scores are consistent and comparable across different forms.

Relevant evidence might include

- Administration procedures are consistent
- Specifications for the development of material are clear and followed
- Piloting and standardising of forms including setting cut scores
- Rater training, standardisation and post moderation
- Checks of intra- and inter-rater reliability
- Assessment-reassessment reliability checks

Open and fair

- There is clear information for students on what is required to be successful in the assessment.
- Assessment tasks, topics and raters are as free as possible of bias.
- Clear information for students on complaints and appeals procedures and special consideration.
-

Relevant evidence might include

- Published outlines
- Bias guidelines and training
- Re-mark and re-sit policies
- Student and teacher feedback.
- Moderation reports and follow-up action plans.
- Policies and procedures related to complaints, appeals, special consideration etc.

Secure

- Assessment responses represent each student's own work and are free of cheating and plagiarism
- Students have not had the advantage of having previously seen assessment material

Relevant evidence might include

- A policy on the control of live assessment material which prevents students gaining access that gives them an unfair advantage.
- Use of plagiarism detection software
- Balance of tasks under full exam conditions.
- Academic misconduct policies and procedures
- Assessment management policies. E.g invigilation, ID check etc
- Policies for storage of assessment materials and completed assessments

Format for applications

Applicants submit file with:

- Overall course and assessment design statement
- Summary chart of claims and references to supporting evidence (see example below)
- High level course outline
- All current assessment tasks
- Assessment design policy, procedures and specifications
- Validation & review records (reliability statistics, student surveys, benchmarking summaries, moderation reports, etc)
- Summary of key staff quals and experience

Valid

Claim

- The assessment can be interpreted as representing ability to cope with the communication demands of English-medium university study at the appropriate level
- Assessment tasks, including the conditions under which they are performed, cover a relevant and sufficient sample of those typical at university
- Texts and language represent academic use
- Judgement criteria and scoring accurately reflect the relevant aspects of ability defined in the construct.

Relevant evidence to support claim includes

- A construct definition
- A reference description/analysis of Target Language Use
- A course outline
- Text analysis (vocabulary, lexical density, etc)
- Expert judgement of appropriacy of texts, tasks & rating (e.g. pre-moderation)
- Clear and firm re-submission policy

For specific evidence see

Construct defn p 1
Course outline p2
Overall assessment design outline p3
TLU description p7
Text analysis stats summary p11
Etc.
Rating scales for each assessment task

Useful reading

Bachman, L. & A Palmer. 2010. *Language Assessment in Practice*. Oxford. OUP.

Knoch, U. & C. Elder. 2013. "A framework for validating post-entry language assessments (PELAs)." *Papers in Language Testing and Assessment Vol. 2, Issue2, 2013*.

Messick, S. 1994. "Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning." ETS Research Report. Princeton, N.J.