

Using Messick's Framework to Validate Assessment Tasks in Online Environments: A Course in Writing Effectively for UNHCR

Valerie Ruhe, PhD., University of Maine

Bruno D. Zumbo, University of British Columbia

**Paper Presented at 2005 American Educational
Research Association conference, Montreal.**

Messick's (1988) Vision for Technology and Assessment

- Technology-based delivery methods will transform teaching, learning and student assessment
- Classical “unitary” approaches to test validity are unsuited to technology-based contexts (Messick, 1988).
- Validity theory is comprehensive BUT validation practice is unitary and fragmented.
- Technology will render this “persistent disjunction” between validity theory and validation practice “no longer tenable”, thereby unifying theory and practice.

The Practical Application of Messick's Vision

- The value of technology-based assessment tasks goes beyond validity coefficients to include diverse aspects of value, such as learner satisfaction, cost-benefit, underlying values and unintended consequences.
- Messick's (1989) four-faceted framework of validity provides a conceptual guide for conducting a comprehensive assessment of merit and worth which includes all of these aspects of value.

Our Purpose

- The debate on values and consequences = Popham (1997) to Shepard (1997).
- Values and consequences are always present in testing, but hidden in the background.
- We will discuss validity and apply Messick's framework to the evaluation data from *A Course in Writing Effectively for UNHCR*.
- Our purpose is to use Messick's framework to illuminate the shadows, bringing values and consequences into the foreground.

The Adapted Messick's (1989) Framework

- Evidence: feedback, grading, completion rates, and learner satisfaction.
- Relevance: task authenticity and cost-benefit analysis e.g. economies of scale.
- Values, theory and ideology.
- Unintended instructional and social consequences
- Validity is a progressive matrix; these aspects of value are dynamic and overlapping.

A Course in Writing Effectively for UNHCR

- A technical writing course for native and non-native speakers of English
- Developed by Commonwealth of Learning (COL) in Vancouver, BC in April, 2000.
- Print/email to over 750 UNHCR employees in 70 countries and 10 time zones.
- 100 learners per cohort, 3 cohorts per year.
- 3 course modules. Assessment tasks: field reports.

Our Methodology (Ruhe, 2002)

- Interviews were conducted with eight learners, two course administrators and five tutors.
- Learner satisfaction survey: 116 respondents from December, 2000 to September 2001.
- Summarized tutor quality assurance scores.

Our Findings: Evidence on Learner Satisfaction

- 100% said the assignments helped them to improve their writing skills
- 91 percent said the feedback was helpful all or almost all of the time
- Interviews: learners appreciated the high standards of the course manual and the lessons on UHHCR terminology

Evidence on Completion Rates

- 2001 course completion rates averaged around 76% percent
- war, floods, pregnancy, illness, surgery, electricity black-outs, abrupt and overwhelming changes in workload and job transfers, bombs, horseback into Pakistani Internet cafés.
- Monthly Progress Report (MPR) increased completion rates

Evidence on Quality Assurance

- Tutors were evaluated on their feedback.
- All TMAs filed with the course administrator.
- Two randomly selected TMAs were scored.
- Tutor rankings assign future contracts.
- March-June 2001: Mean TMA score for all tutors improved from 6.6 to 7.1
- This process reduced variability and “standardized” tutor feedback.

Evidence: Beyond Psychometrics

- Use inter-rater reliability coefficients to compare feedback and grades across tutors.
- COL *extended* this traditional conception of inter-rater reliability into a quality assurance mechanism called the Tutor Marked Assignment (TMA), which actually *reduced* rater error and *standardized* tutor feedback.

Relevance

- Field reports were topics of local interest assigned by supervisors, eg HIV/AIDS and Family Planning Project in central, new refugee camps in western Africa, elementary school education for girls in refugee camps Iraq.
- Authentic, UNHCR letterheads, sample field reports in the UNHCR course manual.
- Language and content were integrated (Mohan and Beckett, 2003).

Cost/benefit

- The course developed to reduce costs.
- Online delivery cheaper than flying learners to face-to-face classes.
- “Up-front” investment vs. lower long-run operating costs e.g. paper, postage, shipping
- horseback rides across borders

Values

- “Tension” between Geneva standard, and 40 different languages from Eastern Europe, Africa, Asia, and South America
- Learners were assessed relative to where they began, not relative to each other. Focus of feedback was “macro” level, e.g. organization, eliminating wordiness, coherence.
- This approach to feedback and grading reflected pluralistic values which honored diversity.

- Some learners asked to be evaluated against *local* not Geneva standards.
- COL's educational methods were very different from those of Myanmar. said that his country, which was an education in itself.
- Assignment deadlines vs. African culture, which placed family obligations above work obligations.
- Deadlines were “euro-centric” and “racist”.
- Tutors balanced pluralistic values with “standards”.

More Value Implications

- Technology-based assessment tasks are pluralistic, adaptive and individualized e.g.
- performance-based checklists for online discussions (Marttunen, 1997).
- Internet-based strategies (Collis, 1998)
- Online summary statistics to track students' content coverage (Harasim, et al., 1996),
- New pluralistic assessment tasks based on multi-media (Baumgartner, 1999).

Unintended Consequences

- Instructional vs. social
- Non-response e.g. “hub” e-mail addresses, slow and unreliable line connections
- The design phase *anticipated* unintended, and built mechanisms to *minimize* them e.g.
- Tutor welcome letters, biweekly reminder letters, the Monthly Progress Reports, the “buddy system” and the Tutor Marked Assignment reviews.

Conclusion

- The validation of online assessment tasks is a rhetorical art, an “argument-based approach” (Cronbach, 1982).
- Using Messick’s framework to guide validation practice provides more evidence than classical approaches by bringing values and consequences into the foreground.
- This new application of Messick’s framework in technology-based contexts is an emerging practice (Ruhe, 2002; Bunderson, 2003; Chapelle et al., 2003)