



EALTA's Guidelines for Good Practice: a test of implementation

J. Charles Alderson & Jayanti Banerjee

Outline

- Background
 - The EALTA Guidelines for Good Practice in Language Testing and Assessment
 - The 'ELPAC' validation study
- Applying the EALTA Guidelines
- Issues arising
- Taking the EALTA Guidelines forward

Central concerns

- Promotion of good testing principles and practices
- Use and application of Guidelines for Good Practice
- Monitoring compliance with good practice standards

The EALTA GGP: background

- General principles:
 - Respect for the students/examinees
 - Responsibility
 - Fairness
 - Reliability
 - Validity
 - Collaboration among the parties involved
- Adopted on 20 May 2006
- Translated into 32 different languages

The EALTA GGP: structure

- Available at:

<http://www.ealta.eu.org/guidelines.htm>

- Three audiences i.e. those involved in:
 - the training of teachers in testing and assessment
 - classroom testing and assessment
 - the development of tests in national or institutional testing units or centres
- One section for each audience

The EALTA GGP: extracts

Section C: Considerations for test development in national or institutional testing units or centres

- How clearly are test purposes specified?
- Do test developers and item writers have experience of teaching at the level the assessment is aimed at?
- What quality control procedures are applied?
- How often are the tests reviewed or revised?

The 'ELPAC' test

- ICAO directive re language proficiency
- Funded and developed by EUROCONTROL:
European Organisation for the Safety of Air
Navigation:
<http://www.eurocontrol.int/index1.html>
- Two papers:
 - Paper 1: Listening Comprehension
 - Paper 2: Oral Interaction
- <http://elpacsample.info/>

Phase One:

The 'ELPAC' validation study

- Proliferation of Aviation English language proficiency tests
- Aviation English tests = high stakes
- Lancaster University - external validation of the English Language Proficiency Test for Air Traffic Controllers (ELPAC)

Phase One: The 'ELPAC' validation study

Focus of the review:

- | | |
|---|--|
| <ul style="list-style-type: none">■ Test specifications■ Content coverage■ Mode of delivery (particularly Paper One)■ Paper setting■ Training documentation■ Administration guidelines■ Examiner training and accreditation | <ul style="list-style-type: none">■ Analysis of trial data■ Feedback from stakeholders■ Standard setting and pass marks■ Parallel forms and test equivalence■ Sustainability of the system■ On-going monitoring |
|---|--|

Phase One: The 'ELPAC' validation study

- Executive Summary
 - Dual purpose:
 - Pull together details of the subsequent report
 - Inform and educate naïve stakeholders
- Format (as per EALTA GGP: Section C)
 - Test purpose and specification
 - Test design and item writing
 - Quality control and test analyses
 - Test administration
 - Test review plans and procedures
 - Washback

Using the EALTA GGP

- Very useful for focusing attention on key aspects of test validity.
- Helped us to synthesise our analysis.
- Helped us to frame key recommendations.

BUT

- Some of the questions were very broad and/or demanded interpretation.
- A number of the questions pertained to a live test rather than a test that was still in development.

Phase Two: The AET* Survey

- ELPAC Validation showed proliferation of English language proficiency tests with little or no evidence on websites of test quality
- Further search to update findings confirmed first survey
- Decision to turn EALTA Guidelines into a survey instrument

Taking the EALTA Guidelines forward

A framework for collecting validity data

- Questionnaire
 - Iterative process
 - Starting point the EALTA GGP – Draft 1
 - Comparison with questionnaire in Alderson *et al* (1995)
 - Lancaster Language Testing Research Group – Three rounds of reviewing
 - Fuzzy concepts in GGP identified
 - Areas where claims would need to be substantiated

Applying the EALTA Guidelines

■ Example 1

- How clearly is/ are test purposes(s) specified?
- Problem: how can respondents define “clear”?

■ Example 2

- What training do test developers and item writers have?
- Problem: What counts as evidence? Number of hours, days, years? Content of training?

Taking the EALTA Guidelines forward

■ Concerns:

- Length
- Specificity
- Supporting documentation
- Maximising response rates

Concern: Length

- Initial solution:

Emphasise in covering letter the crucial importance of such high stakes testing and need for detailed quality control

- Final solution:

Devise an initial shorter questionnaire, and ask about willingness to complete a follow-up questionnaire

Concern: Specificity

- Initial solution:

Adjust GGP to AET context.

- e.g. Reference to ICAO scales, instead of CEFR
- e.g. Remove reference to washback
- e.g. Add questions about needs analysis:
 - AET Q2.8 *Have you conducted an investigation of test-takers' needs?*
 - AET Q2.9 *Have you carried out a survey of how language is used in the context of aviation?*
- e.g. Add questions about computer-based rating of subjective tests

Concern: Supporting documentation

- Initial solution:

Require attachment of supporting evidence or precise URL

- “According to the specifications, what is the purpose of the test? Please attach the statement of test purpose (give the filename here) or give the precise website address (URL) where it can be located.”

- Final solution:

Mainly Y / N questions but a request for follow-up documentation

Concern: Maximising response rates

- Initial solution:

Assurance of anonymity, of importance of responses

- Final solution:

A two-part questionnaire, where the first “filter” did not require technical information, and where opinions were also sought on the ICAO scales, about which many individuals and organisations reportedly have concerns and opinions.

Concern: Maximising response rates

- Final solution (cont):
 - Web-based versions of the filter and follow-up questionnaires
 - Yes/ No questions
 - Occasional open-ended boxes
 - Option for a Word document, or a telephone/ Skype call

Changes/clarifications to GGP

■ Glossing terms:

- Specifications = test blueprint
- Stakeholders = those who commissioned the test or have an interest in it
- Construct = the skills and sub-skills that the test / subtest(s) are intended to measure

■ Deleting questions:

- *What quality control procedures are applied?*
 - (redundant)
- *How is potential test misuse addressed?*
 - (highly unlikely?)

Changes/clarifications to GGP

■ Specifying

- *GGP Q 1.5 Are the specifications for the various audiences differentiated?*
- *AET Q 1.4 Which of the following groups of test-takers are specified?*
 - *Air Traffic Controllers*
 - *Pilots*
 - *Air Service Personnel*
- *Please add any groups your test caters for but which we have not listed.*

Changes/clarifications to GGP

■ Clarifying

- *GGP 4.1 What are the security arrangements?*
- *AET Q 5.1 How is cheating (for example, personation, bribery, copying, access to illicit materials, etc) checked or prevented?*

■ Making relevant

- *GGP Q 5.3 What procedures are in place to ensure that the test keeps pace with changes in the curriculum?*
- *AET 6.2 What procedures are in place to ensure that the test keeps pace with changes in the ICAO requirements?*

Changes/clarifications to GGP

- Adding items
 - e.g needs analysis
- The GGP makes no reference to pass/fail distinctions or grade boundaries, whereas in the case of the Aviation English tests, pass/fail boundaries are clearly crucial,
 - AET Q3.9 *How is the pass mark for objectively scored tests determined?*

Changes/clarifications to GGP

- Adding items (cont/..)
 - AET Q 5.5: *What processes are in place for test-takers to make complaints or seek scrutiny of marks, and/or re-marking of the test?*
 - AET Q 5.6: *How often can a candidate re-take the test?*

Changes/clarifications to GGP

- Modifying double barrelled questions
 - GGP Q 1.10 *Are marking schemes/rating criteria described?*
 - AET Q 1.8 *Are the rating criteria available?*

- Changing ambiguous wording
 - GGP Q 3.2 *Are the tests piloted?*
 - AET Q 4.1 *Are the tests trialled?*

 - GGP Q 3.3 *What is the normal size of the pilot sample....*
 - AET Q 4.2a. *What is the normal size of the trial sample?*

Evidence of claims

- At the end of the questionnaire, we offered respondents the opportunity to send us further details by email, and provided a list of documents or information that we would find it very helpful to have access to.

Implications for the EALTA Guidelines

- The Guidelines are intended for consciousness-raising
- The Guidelines need amending to become a survey instrument in order to elicit more specific information
- A big problem is practicality: the GGP cover many aspects of quality control, but a survey instrument cannot be comprehensive
- Response rates poor: how to maximise?

Looking beyond the EALTA Guidelines

- How can organisations be “obliged” to respond?
- Naming those who do not respond?
- Identifying good practice by name?
- Identifying bad practice by implication, i.e. by listing all those who DID respond but not identifying bad practice by name?
- In a field as crucial to safety as Aviation Language testing, why is there no organisation formally responsible for ensuring test quality?