

Auditing the Quality Profile

From Code of Practice to Standards

Introduction

This paper reports on the work of the Association of Language Testers in Europe (ALTE) and in particular the *ALTE Procedures for Auditing* and the outcomes of audits conducted in 2007 and 2008. It summarises the background to the approach, presents data collected so far and reflects on the lessons to be learnt in implementing this approach. The need to justify the use of language assessments to stakeholders using codes of practice and quality management systems forms part of this discussion.

The *ALTE Procedures for Auditing* derive from ongoing work on the *ALTE Code of Practice* (1994) and draws on approaches to auditing adopted by other organisations such as ISO – the International Standards Organisation.

The formal scrutiny of standards is the culmination of a long process of working towards establishing audited “quality profiles” across the ALTE examinations. The aim of the process is to allow ALTE members to make a formal, ratified claim that a particular test or suite of tests has a quality profile appropriate to the context and use of the test based on 17 parameters for establishing minimum standards. Ultimately this is to ensure that the assessment is fair and meets the needs of the stakeholders in appropriate ways.

It is important to remember the context in which this work has been carried out and in particular the wide range and diversity of ALTE members, associates and affiliates:

- 31 full and associate members: these are organisations which include government departments, universities, consortia and examination boards which have a role in assessing their own language.
- 26 languages are currently represented, including many less widely taught languages.
- In addition, there are nearly 40 institutional affiliates with an interest in language education and language assessment.

The description of the development and application of this system ties in with recent discussions in the language testing literature on the use of argumentation to support claims of validity. The system is also an example of another kind of theory, “a theory of action” (or change processes – see Fullan: 1993, 1999).

A theory of action, capable of bringing about positive and sustainable change, was needed by ALTE members in order to raise standards and improve the quality of their examinations. I would argue that the QMS approach and the auditing system have now provided such a theory.

Ethics and Principles of Good Practice

The early work of ALTE Members in this area took place in the 1990s when they addressed the question of what a code of practice might be like and what the underlying principles should be.

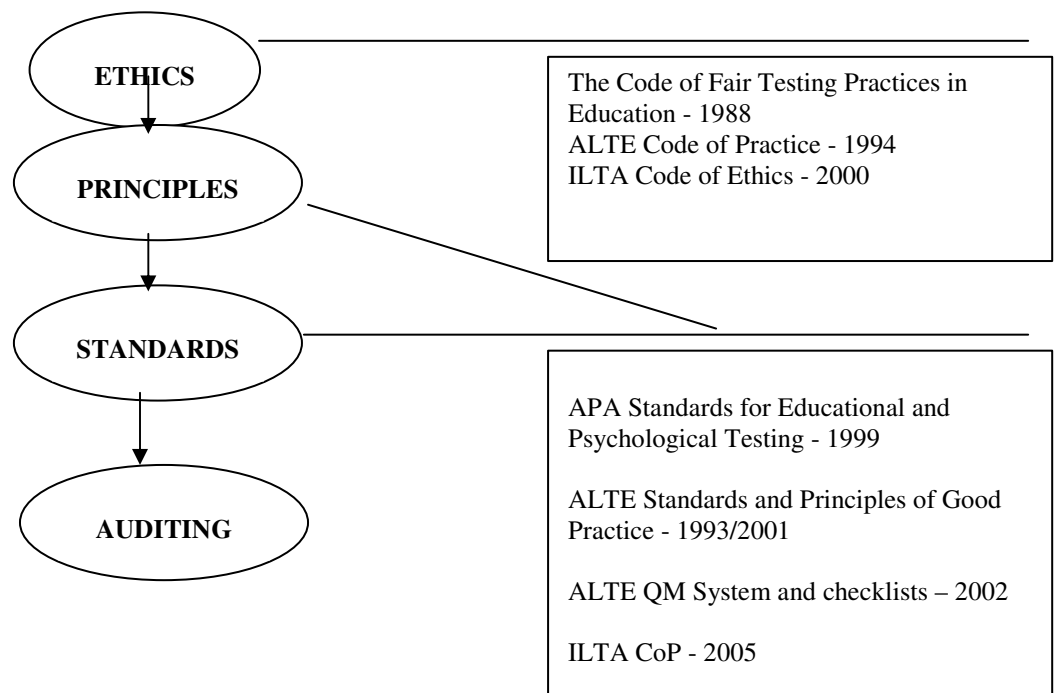
This led to the publication of the ALTE *Code of Practice* in 1994, and at about the same time, the adoption of the ALTE *Standards and Principles of Good Practice*.

A Code of Practice Working Group was set up in 2000 to take the early work forward in light of changes to the ALTE membership and advances in the field of language testing in the 1990s.

Since then this group has met regularly and the main outcomes include the production of detailed documentation up-dating the Principles of Good Practice, proposals for minimum professional standards and the implementation of appropriate Checklists and procedures for monitoring those standards. The most recent work starting about 3 years ago has focused on the development of the ALTE auditing system.

Methodological framework

The methodological framework which this work represents can be shown as follows, with some relevant reference documents listed on the right:



The overriding aim can be summarised as follows:

“..... to strive for overall fairness in the testing process, always acting in good faith and avoiding negative impact, in order to provide a high quality of service to the clients and other test users”.

However, the ALTE Code was an aspirational document and not designed to assist language testing practitioners in carrying out their day-to-day work of writing and administering tests, or in agreeing on what might be acceptable in terms of minimum standards for their work.

Having established the principles, and having provided some practical tools to help ALTE Members improve their examination systems, the working group addressed the issue of *how to put the principles into practice*, how improvements could be monitored and whether adequate standards were in fact being met. In order address this problem and to seek consensus, it was decided that the appropriate paradigm for this activity would be that of *Quality Management Systems*.

A Quality Management System (QMS)

QM systems seek to improve the products and/or services of an organisation in order to meet the requirements of its customers in the most effective way, and they go about doing so in a well-planned and focused way. There are many examples of QM systems being used in educational contexts, and several of these were thoroughly reviewed by the working group. This was summarised and then extensively discussed by the full membership (for more details see van Avermaet et al 2004).

Effective QM Systems usually involve a public statement in the form of a Code of Practice or Charter, and an ongoing commitment to the *change process* (improvement) typically involving the following steps:

- Define your mission, role of institution, future ambitions/aims
- Assess what you currently do
- Identify areas in need of improvement
- Decide on measures of improvement and an action plan
- Carry out action to bring about improvement
- Review progress and revise the action plan
- Carry out action to bring about improvement
- *And so on*

The ALTE Code set out the public position in terms of the aspirations and ambitions of the association, as explained above, but in adopting the QMS approach, Members undertook:

- a) to understand the nature of their organisations better;

- b) to involve their stakeholders in making improvements to the quality of their examination systems.

In effect, this involved an ongoing commitment to “change management” (innovation processes). It is axiomatic in this approach that improvement is always possible, even where good practice may already exist.

In QM systems standards are not imposed from "outside", but are initially established through the system itself and the procedures to monitor standards are based on *awareness raising* and *self-assessment* in the first instance. External monitoring is introduced at a later stage to confirm that the minimum standards are being met.

In order to provide a practical tool to raise awareness of those areas where change was necessary the original Code of Practice was reworked to function as *Quality Management Checklists*; this re-designed format reflected the four aspects of *testing cycle* with which test developers are very familiar:

- Examination development
- Administration of the examinations
- Processing of the examinations including the marking, grading and issue of results
- Analysis and post-examination review

The revised format provided four checklists which were put into Excel spreadsheets for ease of use and for Members to use as evaluation tools. These Excel checklists are available from the ALTE Website (www.alte.org).

In summary, the QMS approach provided a supportive tool and allowed those working with it:

- a) to enhance the quality of their examinations in the perspective of fairness for the candidates;
- b) to engage in negotiations with their senior management and sponsors in a process of organisational change where necessary, and to ensure that resources are made available to support on-going improvements;
- c) to move from self evaluation to external verification in order to agree and monitor acceptable standards.

By proceeding in this way ALTE Members were made constantly aware of the different contexts in which they work and of the various backgrounds from which the different Members come. Extensive discussions have taken place around the question of how to achieve a reconciliation between diversity and professional standards which are acceptable to all. Since 2007 this has continued through the implementation of the auditing system.

Monitoring standards – Auditing the Quality Profile

It was always envisaged that self-evaluation would need to be supplemented by an external “auditing” system. This was developed and piloted starting in 2005/6 and in 2007/8 we have seen the culmination of the process of establishing audited “quality profiles” across a wide range of ALTE examinations.

Taking the Code of Practice and QMS checklists into account, seventeen parameters for establishing *minimum standards* were agreed with the aim of establishing a *Quality Profile* for each exam or suite of exams (see Appendix 2). The formal external scrutiny of these parameters in the auditing process is intended to ensure that adequate standards are being set and achieved.

ALTE Members are required to make a formal, ratified claim that a particular test or suite of tests has a quality profile appropriate to the context and use of the test bearing in mind the following points:

- Different tests are used in different contexts, by different groups of test users. There is no intention to impose a single set of uniform quality standards across for all ALTE exams.
- Members requesting an audit of their quality systems and procedures are invited to *build an argument* that the quality standards within a test or suite of tests are sufficient and appropriate for that test or suite of tests.
- It is the *argument* which is the subject of the audit, rather than the organisation itself (which is often dealt with by other systems of regulation, e.g. ISO 9001, government regulators etc.)
- Each audit considers one test, suite of tests or testing system.
- The audit has both a consultancy and quality control role.
- The audit aims to establish that minimum quality standards are being met in a way that is appropriate to the context of a test, and also to offer recommendations towards best practice where, though quality standards are appropriate, there is still room for improvement.
- If quality standards are not being met, ALTE Members will collaborate with the audited organisation to implement an action plan aimed at working towards and ultimately reaching the quality standards.

The argumentation structure is as follows.

A claim is made about each of the seventeen parameters; the claims support an argument that minimum standards are being met for the test in question. Information is provided to support the claims; this information is provided in the form of explanations. A justification is also needed to provide legitimacy for this information; this must be based on the relevant language testing theory with reference to the Code of Practice and the Principles of Good Practice, and may also take into account prior experience and best practice models where appropriate. This justification, in turn, needs to be backed up with appropriate evidence which has been collected as part of the validation process.

This approach is consistent with Toulmin's (2003) argument structure and Bachman's application of his work to language testing (e.g. Bachman 2005). In Toulmin, the justification is known as a *warrant*, and the evidence is known as the *backing*. Bachman suggests that the test developer must be prepared to deal with *rebuttals* (alternative explanations and counter claims) and to provide additional evidence to reject them (*rebuttal data*).

The QM checklists provide a practical tool for structuring the explanations supporting the claims and for collecting and presenting the necessary evidence.

The auditor and an elected Standing Committee scrutinise the claims and the evidence; they can challenge whether the claims adequately meet minimum standards, and if necessary can ask for additional information to be provided. An audit remains "ongoing" or "in progress" until such points have been clarified or alternative procedures have been put in place which are deemed acceptable.

In 2006/7 a Manual setting out the "procedures for auditing" was developed, a small group of auditors was recruited and appointed from within the ALTE membership and an initial induction and training programme was developed. All those eligible to conduct audits had to go through this training process. It was recognised, however, that the procedures would need to be improved and that better coordination and standardisation would be required in future as the system develops (i.e. an ongoing and iterative process in its own right).

The membership as a whole is the arbiter of decisions arising from the auditing process; this takes place through the Council of Members as a whole and in particular through the smaller, elected Standing Committee which has delegated responsibility to oversee the auditing process.

When outcomes of an audit are considered *satisfactory*, but only minimally so for some of the parameters, the Standing Committee can recommend that an *action plan* be developed for that examination in order to improve practice with regard to the weak parameters. Members are asked to provide an action plan with a timescale and a rationale for addressing the issues in a particular way. Even in cases of good practice, *opportunities for improvement* (OFIs) can be identified and suggested.

Results of the auditing

After the initial piloting of the system, further refinements have been made and implemented in the audits carried out in 2007/8.

The first cycle of auditing is still going on; this cycle aims at having at least one examination of each member audited. This first cycle also provides the necessary experience for improving the audit procedures and the training system for the future.

- 30 members expect to have their examinations audited in this cycle plus several new members.
- 32 examinations of 15 members have been audited to date.
- The auditing of the tests of 10 more members is now in progress.
- 4 are still in a negotiation phase for setting up the audit.
- 3 new member audits are in progress (Czech, Trinity College London, CCIP France) and one new member - the Government of Navarra – is waiting to have the audit arranged.

The outcomes of the completed audits (at the time of writing) are as follows:

- 2 remain in progress (not satisfactorily meeting minimum standards yet).
- 7 satisfactory with an action plan.
- 5 satisfactory with suggestions for improvements (OFIs).
- Several reports have been submitted but are still to be discussed by the Standing Committee.

Reflections on the auditing experience

After each audit the auditor has to write an audit report, which is first discussed with the auditee and then sent to the Standing Committee to be discussed and ratified. Despite the fact that all the auditors followed the guidelines as described in the “procedures for auditing” each report differed from the others to a certain degree and in significant ways. This probably had to do with the following issues:

- Differences in background of the auditors, leading to a different focus of attention.
- A need for greater elaboration of the core elements within each minimum parameter to achieve better standardisation of the pre-audit, the audit and the reporting.
- A need for more clarity in and agreement on which core elements have to be met in order to meet the minimum standard for each parameter.

Comparisons of the information in the different audit reports and the way different auditors came to their judgments has already resulted in a more accurate description of the core elements of each minimum parameter. For example better guidance in dealing with Minimum Parameter 13:

“You collect and analyse data on an adequate and representative sample of candidates and can be confident that their achievement is a result of the skills measured in the examination and not influenced by factors like LI, country, gender, age and ethnic origin.”

Or Minimum Parameter 14:

“If you make a claim that the examination is linked to an external reference system (e.g. the Common European Framework) then you can provide evidence of alignment to this system.”

Working in this way it is possible to make audits more comparable and transparent and less dependent on individual interpretations of the different auditees and auditors.

The completed audits have also provided a useful “snap shot” of the state of affairs across the examinations of ALTE members. The information now functions as an input for further training and for organising well targeted workshops to help improve examinations.

Way of improving the auditor training itself have also emerged based on the experiences of the auditing described above. This process of improvement will continue in the future, but so far it has already resulted in an improved training programme, that is more firmly based on ‘real life auditing data’.

Conclusion

In conclusion, the QMS and the auditing procedures provide a dynamic system which forms the necessary basis for action in establishing and monitoring standards;

It is useful for:

- clarifying the quality demands of examinations in relation to their functions and purposes;
- providing ALTE-members valuable information of the state of affairs in the examinations in their frameworks;
- providing ALTE members concrete possibilities or necessities for improving their own examinations;
- setting priorities for training, workshops and mutual consultancy and support;
- accounting for the validity of the examinations to stakeholders;
- improving the system itself.

References:

- AERA/APA/NCME (1985/1999). *Standards for educational and psychological testing*. Washington: AERA.
- ALTE (1993/2001). Principles of Good Practice for ALTE Examinations.
- ALTE (1994). Code of Practice.
- ALTE (1998). Handbook of Language Examinations and Examination Systems.
- ALTE/Council of Europe (1997). Users Guide for Examiners. Strasbourg.
- van Avermaet, P. Kuijper, H. ,and Saville, N. (2004). A Code of Practice and Quality Management System for International Examinations. *Language Assessment Quarterly*, 1(2&3). Special Issue: The Ethics of Language Assessment. Guest Editor, Alan Davies.
- Bachman L.F. (1990). *Fundamental considerations in language testing*. Oxford: OUP.
- Bachman, L.F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34. LEA inc.
- Bachman, L. F. and Palmer, A. (1996). *Language Testing in Practice*. Oxford: OUP.
- Fullan, M. (1993). *Change Forces: Probing the Depths of Educational Reform*. London: the Falmer Press.
- Fullan, M. (1999). *Change Forces: The Sequel*. London: the Falmer Press.
- ILTA (1995). Task Force on Testing Standards.
- ILTA (2000). Code of Ethics. *Language Testing Update*, Issue 27.
- Kane, M.T.1992. An argument-based approach to validity. *Psychological Bulletin*. 112, 527-535.
- Kunnan, A. (2000) (Ed). *Fairness and validation in language assessment*. Cambridge: UCLES/CUP.
- Kunnan, A..J. 2004. Test Fairness. In Milanovic and Weir 2004.
- Messick, S..A. 1989. Validity. In Linn (ed.) *Educational Measurement*. New York: ACE / Macmillan.
- Messick, S. 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 32(2), 13-23
- Messick, S.A. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35.
- Mislevy, R.J. 1994. Evidence and inference in educational assessment. *Psychometrika*, 59, 439-483.
- Mislevy, R.J., L.S. Steinberg, and R.G. Almond. 2003. On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-62.
- Saville, N. Van Avermaet, P. and Kuijper, H. Setting and Monitoring Professional Standards: a QMS approach. Paper presented at the 2nd EALTA Conference, Voss, June 2005.
- Toulmin, S.E. (2003). *The uses of argument*. (updated). Cambridge: CUP.

Appendix 1 – ALTE Membership (2008)

26 languages	31 full members: organisations comprising government departments, universities, consortia and examination boards
Basque	Basque Government
Bulgarian	University of Sofia "St Kliment Ohridski"
Catalan	Generalitat de Catalunya
Danish	Danish Language Testing Consortium
Dutch	CITO Certificaat Nederlands als Vreemde Taal
English	University of Cambridge ESOL Examinations
Estonian	National Examination and Qualification Centre
Finnish	University of Jyväskylä and National Board of Education
French	Alliance Française Centre international d'études pédagogiques
Galician	Xunta de Galicia
German	Goethe-Institut telc GmbH TestDaF-Institut
Greek	University of Athens
Hungarian	Institute of the Eötvös Loránd University
Irish	National University of Ireland
Italian	Università per Stranieri di Perugia
Latvian	Naturalisation Board of Latvia
Lithuanian	University of Vilnius
Luxembourgish	Centre de Langues Luxembourg
Norwegian	University of Bergen
Polish	Jagellonian University
Portuguese	University of Lisbon
Russian	Russian Language Testing Consortium
Slovenian	University of Ljubljana
Spanish	Instituto Cervantes and University of Salamanca
Swedish	Stockholm University
Welsh	Welsh Joint Education Committee

Appendix 2 - Minimum standards (2008)

17 parameters for establishing Quality Profiles in ALTE examinations

The Quality Profile is created in each case, by explaining how the examination meets the following minimum standards, and by providing adequate evidence.

1	Test construction	The examination is based on a theoretical construct, e.g. on a model of communicative competence.
2		You can describe the purpose and context of use of the examination, and the population for which the examination is appropriate.
3		You provide criteria for selection and training of test constructors and expert judgement is involved both in test construction, and in the review and revision of the examinations.
4		Parallel examinations are comparable across different administrations in terms of content, stability, consistency and grade boundaries.
5		If you make a claim that the examination is linked to an external reference system (e.g. Common European Framework), then you can provide evidence of alignment to this system.
6	Administration & Logistics	All centres are selected to administer your examination according to clear, transparent, established procedures, and have access to regulations about how to do so.
7		Examination papers are delivered in excellent condition and by secure means of transport to the authorized examination centres, your examination administration system provides for secure and traceable handling of all examination documents, and confidentiality of all system procedures can be guaranteed.
8		The examination administration system has appropriate support systems (e.g. phone hotline, web services etc).
9		You adequately protect the security and confidentiality of results and certificates, and data relating to them, in line with current data protection legislation, and candidates are informed of their rights to access this data.

10		The examination system provides support for candidates with special needs.
11	Marking & Grading	Marking is sufficiently accurate and reliable for purpose and type of examination.
12		You can document and explain how marking is carried out and reliability estimated, and how data regarding achievement of raters of writing and speaking performances is collected and analysed.
13	Test analysis	You collect and analyse data on an adequate and representative sample of candidates and can be confident that their achievement is a result of the skills measured in the examination and not influenced by factors like L1, country of origin, gender, age and ethnic origin.
14		Item-level data (e.g. for computing the difficulty, discrimination, reliability and standard errors of measurement of the examination) is collected from an adequate sample of candidates and analysed.
15	Communication with Stakeholders	The examination administration system communicates the results of the examinations to candidates and to examination centres (e.g. schools) promptly and clearly.
16		You provide information to stakeholders on the appropriate context, purpose and use of the examination, on its content, and on the overall reliability of the results of the examination.
17		You provide suitable information to stakeholders to help them interpret results and use them appropriately.