

## USE OF QUALITY MODELS AND INDICATORS FOR EVALUATING TEST QUALITY IN AN ESP COURSE

IEVA RUDZINSKA

Ieva.Rudzinska@lspa.lv

Latvian Academy of Sport Education, Latvia



### ABSTRACT

Qualitative methods of assessment play a decisive role in education in general and in language learning in particular. The necessity to perform a qualitative assessment comes from both increased student competition in higher education institutions (HEIs), and hence higher demands for fair assessment, and a growing public awareness on higher education issues, and therefore the need to account for a wider circle of stakeholders, including society as a whole. The aim of the present paper is to study the regulations and laws pertaining to the issue of assessment in Latvian HEIs, as well as to carry out literature sources analysis about assessment in language testing, seeking to select criteria characterizing the quality of English for Specific Purposes (ESP) tests and to apply the model of evaluating the quality of a language test on an example of a test in sport English, developed in a Latvian higher education institution.

An analysis of the regulations and laws about assessment in higher education and literature sources about tests in language courses has enabled the development of a test quality model, consisting of seven intrinsic quality criteria: clarity, adequacy, deep approach, attractiveness, originality/similarity, orientation towards student learning result/process, test scoring objectivity/subjectivity. Quality criteria comprise eleven indicators. The reliability of the given model is evaluated by means of the whole model, its criteria and indicator Cronbach's alphas and point-biserial (item-total) correlations or discrimination indexes DI. The test was taken by 63 participants, all of them 2nd year full time students attending a Latvian higher education institution.

A statistical data analysis was performed with SPSS 17.0. The results show that, although test adequacy and clarity is sufficiently high, attractiveness and deep approach should be improved. Also the reliability of one version of the test is higher than that of the other one. One of the ways to improve test quality could be to involve other HEIs in the process of designing tests, because in a small institution it is difficult to collect authentic material for test design and create reliable language tests in a narrow field (in our case: sport English).

**Key words:** ESP, quality of language tests, test reliability, higher education.

### INTRODUCTION

Assessment plays a key part in education in general, and in language education in particular, because it has a considerable impact on student learning. It is necessary to pay even more attention to the quality of assessment owing to both increased student competition in higher education institutions (HEIs) and an increased society awareness about higher education issues, which entails higher demands for fair assessment.

The aim of this paper is twofold. In the first instance, it seeks to raise awareness about the diversity of criteria, qualitative ESP tests should comply with, studying the regulations and laws pertaining to the issue of assessment in Latvian HEIs, carrying out literature sources analysis about fair assessment in language testing, and finally developing English for Specific Purposes (ESP) test quality model, comprising a list of quality characterizing criteria, consisting of at least several quality indicators. Simultaneously, it also aims to investigate the quality of an ESP Test against the framework of the selected model and with the help of selected criteria.

The methods of research utilized in this paper include:

- conducting state regulation and law review and literature sources analysis about the issue of fair assessment in language testing
- working out a model for evaluating the quality of ESP tests and checking its reliability
- designing a questionnaire to evaluate test quality within the framework of the developed model
- evaluating an ESP test quality within the framework of the developed model.

## THEORETICAL FOUNDATIONS

Language competence is a dynamic combination of professional, communicative and intercultural competences (Luka, 2008, p.152.) Communicative competence implies an effective use of all four language skills which carry out a communicative function. Intercultural competence (Stiers, 2004; Korhonen, 2004) consists of communicative competence, the ability to act in intercultural communication contexts, and international working experience. Professional foreign language competence, developed in an ESP (English for Specific Purposes) course, is a combination of communicative and intercultural competences, as well as professional competence, whose inseparable part constitutes professional experience.

Classical test qualities are validity, reliability and practicality. Valid tests assess what they are designed to assess, and reliable tests do it in a systematic way. Contemporary ESP tests assess the use of language for specific purposes and competences in situations that should resemble real-life situations as close as possible.

However the mentioned test quality indicators do not meet all the needs of contemporary assessment. Thus Latvian higher education laws and regulations indicate that assessment in education should encompass such factors as (1) summing up positive achievements and reflecting the student's development; (2) fairness, including openness and clarity of assessment criteria; (3) various forms of assessment and adequacy: test correspondence to the content of the study course, knowledge acquired and skills and competences developed (MK noteikumi Nr. 141 "Noteikumi par valsts pirmā līmeņa profesionālās augstākās izglītības standartu, MK noteikumiem Nr.347 "Noteikumi par valsts pirmā līmeņa profesionālās augstākās izglītības standartu"; Ministru kabineta noteikumi Nr.2 „Noteikumi par valsts akadēmiskās izglītības standartu”; Ministru kabineta noteikumi Nr.481 “Noteikumi par otrā līmeņa profesionālās augstākās izglītības valsts standartu”).

One of the recent developments in the field of foreign language testing in higher education institutions in the EU are GULT (Guidelines for University Language Testing) task-based tests (GULT Project description, on-line), which use authentic reading and listening materials. Such tests can be developed only through cooperation of several HEIs that specialize in the field under consideration.

Quality assurance is usually applied on a HEI or study program level, not on one study course level. To instill quality assurance ideas in separate study courses, enabling separate lecturers more active participation in quality assurance process in their study courses and promoting bottom-up approach, European researchers have developed several models for evaluating the quality in one study course (Lansnier, 2007; Meder, Iske, 2009; Rudzinska, 2009). The quality of a study course usually is evaluated in several blocks (such as, for example, objectives, didactic methods, student cognition processes and cooperation, assessment, results, etc.), and according to a list of criteria, such as clarity, adequacy, deep approach, attractiveness, etc. The quality model developed by Rudzinska (2011), for example, consists of six blocks and six criteria: adequacy, clarity, attractiveness, deep approach, individual work, cooperation. Although assessment is an integral part of the models mentioned above, in the block of assessment there should be included some additional criteria: control works need to be scored both in an objective and subjective way; they should be both original and similar to other; they should assess not only study results, but also study process (Meder, Iske, 2009).

Test adequacy means that test tasks simulate the use of language and tasks which test takers would actually perform in real life situations. Test clarity means that tasks and scoring criteria are clear and unambiguous. Test attractiveness is connected with interactivity and variety (different test tasks, different skills and competences being tested, a possibility to choose from several authentic materials).

The inclusion of the criterion of deep approach in test quality model reinforces one of the main aims of higher education, namely, to encourage students to use higher cognitive processes and promote long-term learning for longer term (Dominowski, 2002; Biggs, 2003). If considered from the bottom up (lower to higher), the main focus of cognition is the ability to remember, understand, apply, analyze, evaluate and create (Bloom, 1992; Anderson, Kratwohl, 2001). Although it is usually mainly lower level cognitive skills that are used for solving tests, almost all test tasks can also test higher cognitive processes (Dominowski, 2002). Each quality criterion - clarity, adequacy, deep approach, attractiveness, originality/similarity, orientation towards student learning result/process, test scoring objectivity/subjectivity - is evaluated with the help of several (two-four) quality indicators.

The reliability of the developed test quality model mainly concerns quality model construct validity. Cronbach alpha values characterize the inner consistency of a test quality model, and of test quality criteria and indicator scales. Both should be higher than 0.8. The correlations between different criteria should be fairly low: from 0.3 to 0.5 because different criteria characterize different aspects of test quality. Component (quality criterion or indicator) correlations

with the whole model characterize a higher level of order, therefore they should be higher, possibly around 0.7 (Alderson, Clapham & Wall, 1995, p.184). The latter correlations are calculated as point-biserial correlations (item-total correlations in SPSS program).

## RESEARCH PARTICIPANTS

The study included 63 2<sup>nd</sup> year students of Latvian Academy of Sport Pedagogy who took a test in sports English (30 of these students did a questionnaire about the quality of the test). The sample of 30 students was a convenience sample, which, however, included the most characteristic cases (Geske, Grīnfelds 2006, p.184): students from all groups in Year 2, as well as a proportional number of women and men (15).

## QUESTIONNAIRE

The questionnaire, which was designed to enquire the students about their opinions about test quality, consisted of 9 questions. Two quality indicators characterized test clarity (CLA): "Test tasks are clearly formulated," "Assessment criteria are clearly formulated"; three of them - test adequacy (ADE): "Test tasks are connected with the aims of the study course," "Test tasks correspond to my level of English," "Test tasks correspond to language learning activities, practiced in the course"; and final three - test attractiveness (ATT): "The materials used in the test come from real life situations and authentic sources," "The topics and problems used in the test are the kind of thing that I can deal with in real life." "Test tasks are varied."

Answers were provided on a scale from 1 to 4, 5th choice being: N/A: not applicable. The respondents had to evaluate whether the test is objective/subjective, original/similar to others, oriented toward study process/result. They could choose from 5 options, from objective/subjective and other abovementioned continuums.

## DATA ANALYSIS

Statistical analysis of the data has been performed with SPSS 17.0 software. Test quality along the criteria of ADE, CLA and ATT is calculated as median values of quality indicators. Wilcoxon Signed Ranks Test is used to identify statistically significant differences between test criteria and their indicators.

Evaluation of test quality, along the criterion of deep approach, is carried out with the aid of test task qualitative analysis, which allows implying what cognitive processes are involved in performing a test task.

Quality model reliability or the reliability of the results, obtained with the developed test quality model, is calculated using Cronbach alpha values of the whole test, quality criteria and indicators, as well as item-total correlations (D.I.) between separate components (criteria and indicators) and the whole model.

## RESULTS

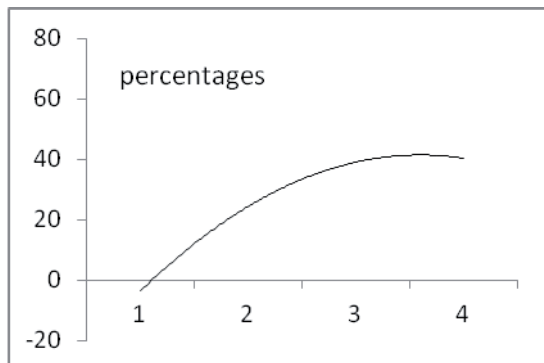
The results show that the designed model is reliable for test quality evaluation, and they give some insight into the quality of the test. Reliability analysis of test quality model reveals that Cronbach's alpha of the developed Test quality model is 0.88. Thus, it is higher than the acceptable value (0.80).

The analysis of the quality of the Sport English test with the help of the developed model revealed that Cronbach's alpha of separate quality criteria were high enough for adequacy and clarity criteria (from 0.73 to 0.76), and not high enough for attractiveness criterion (0.52). Discrimination indexes D.I. (item-total or point-biserial correlations) were acceptable for adequacy and clarity criteria (0.80) and not acceptable for attractiveness criterion (0.74).

### DESCRIPTIVE STATISTICS FOR TEST QUALITY CRITERIA AND INDICATORS

Wilcoxon Signed Rank Test has revealed that, according to the respondents, the clarity, attractiveness and adequacy of the test were developed to the same extent. The students had evaluated the quality of the test as high: median values for all quality indicators are from 3 to 4 (Figure1).

**Figure 1.** Distribution of student answers to the question: "Test tasks are sufficiently varied?" (1 - totally disagree, 4 - fully agree). IR.



### THE FULFILLMENT OF THE CRITERION OF DEEP APPROACH

Table 1 summarizes cognitive activities the students might have used while doing test Task 1, Task 2 and Task 3.

Task 1 of the ESP test (in Sport English) is a translation task. Test takers have to translate from English into Latvian a passage, which in detail describes the techniques of the performance of an exercise in gymnastics.

Task 2 is a production task. Using pictures as stimuli test takers have to describe in sport English, how to perform a stunt (cartwheel, handstand, a.o.) in gymnastics.

Task 3 is a Use of English task or a grammar task, concentrating on the use of participles in texts about gymnastics and in sports texts in general. Test takers have to translate from English into Latvian separate sentences with participles and identify their forms.

**Table 1.**  
Cognitive activities used in doing Task 1 and Task 3

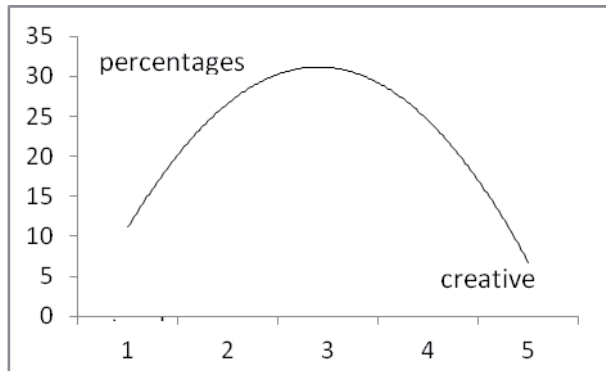
Task No.	Cognitive activities	Level of cognitive activities
1	Recall from memory: 1) translation of specific terms, e.g., <i>workout, lower back, arching, quadriceps, reps, set of exercises</i> 2) translation of general English words, e.g., <i>against, angle, apart, squat, fold.</i> Remember: form of simple future tense (won't move)	R (remembering)
	Interpret, infer, explain (the execution of the exercise)	U (understanding)
	Apply rules of grammar (word-building) to translate verb "strengthen" and noun „width"	Ap (application)
	Analyze (how specific movements relate to the whole exercise)	An (analysis)
	Evaluate (possibility to execute the exercise described)	Ev (evaluation)
	Create new text in another language (process of translation)	C (creation)
2	Recall from memory: 1) translation of specific terms, 2) translation of general English words	R (remembering)
	Interpret, infer, explain (the execution of the stunt)	U (understanding)
	Analyze (how specific movements relate to the whole exercise)	An (analysis)
	Evaluate (possibility to execute the exercise described)	Ev (evaluation)
	Create new text in another language	C (creation)
3	Recall from memory and identify participles and their forms in English and Latvian	R (remembering)
	Produce grammatically and semantically correct translation from English into Latvian	Ap (application)

IR.

Table 1 shows that low and medium level cognitive activities are used more often than high level ones. To perform Task 1, the students have to activate all level cognitive activities, but while performing Task 3, they are supposed only to remember and to apply grammar and word-building procedures in standard situations.

## DESCRIPTIVE STATISTICS FOR QUALITY CRITERIA, BEING EVALUATED ON A CONTINUUM

Median values for quality criteria, being evaluated on a continuum, are from 2 to 3. This result means that the examined test is both standard and creative (Figure 2), and process and result-oriented. However, its scoring is more objective than subjective.



**Figure 2.** Distribution of student answers on the continuum "Test tasks are standard (1) to creative (5). IR.

## CONCLUSION AND DISCUSSION

To meet the demands of contemporary society for fair qualitative assessment, it was necessary to raise awareness about the diversity of criteria, which characterize a qualitative ESP test. Laws, regulations and literature sources analysis enabled the development of test quality model, comprising seven quality criteria - clarity, adequacy, deep approach, attractiveness, originality/similarity, orientation towards student learning result/process, test scoring objectivity/subjectivity. Quality model reliability analysis confirmed that the developed model can be used as a reliable framework for evaluating ESP test quality. The inner consistency for evaluating the criteria of adequacy and clarity is high enough, but it is insufficient for evaluating the criterion of attractiveness. Therefore, in order to increase the reliability of the evaluation of test quality, there should be added more indicators that could characterize attractiveness.

The developed test quality model was applied for the evaluation of an ESP test in Sport English. Wilcoxon Signed Rank Test has revealed that clarity, attractiveness and adequacy of the examined test are equally developed. As regards its compliance with the quality criteria, which are evaluated on the continuum, it can be concluded that the test is both standard and creative one, as well as learning process and result oriented. Test scoring, however, is more objective than subjective. To ensure balance between opposite qualities, assessing students on an ESP course, besides tests should be used other forms of control works (presentations, projects, discussions, etc.), the scoring of which is more subjective.

Deep approach in the ESP Test, which was used as an example, showing the possibilities of the application of the developed test quality model, is realized only partly because low and medium level cognitive activities are used more than the

high level ones. To promote deeper approach, Task 3 (grammar task) could not be presented as separate task, but be incorporated in Tasks 1 and 2.

Another way of rectifying the test is to use authentic reading and listening materials, as practiced in GULT tests. The development of such tests is difficult, especially when carried out by individual lecturers. However, it is our belief that the concerted efforts of staff members, or even higher education institution, could result in the preparation of high quality ESP tests, which embrace all the diversity of quality criteria characterizing a qualitative ESP test.

## REFERENCES

- Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Anderson, L.W., Kratwohl, D.R. (eds.) (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives*. New York: Longman.
- Bachman, L.F., Palmer, A.S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Biggs, J. (2003). *Teaching for Quality Learning at University*. Maidenhead: Open University Press.
- Bloom, B.S. (1992). *Taxonomy of Educational Objectives, Cognitive Domain*. Longman.
- Dominowski, R. (2002). *Teaching Undergraduates*. London: LEA Publishing.
- Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.
- Geske, A., & Grīnfelds, A. (2006). *Izglītības pētniecība: mācību grāmata augstskolu izglītības un pedagoģijas profesionālo un akadēmisko studiju programmu studentiem*. Rīga: LU Akadēmiskais apgāds.
- GULT (*Guidelines for University language Testing*) Project description. Retrieved May 15, 2010, from <http://gult.ecml.at>.
- Korhonen, K. (n.d). *Developing Intercultural Competence as a Part of Professional Qualifications*. *Journal of Intercultural Communication*, 7, pp.1-8. Retrieved October 12, 2007, from <http://www.immi.se/intercultural>.
- Lasnier, J.C. (2003). *Quality Version 2003*, Retrieved from <http://www.quiltnetwork.org>.
- Luka, I. (2008). *Profesionālās angļu valodas kompetences veidošanās augstskolā, Monogrāfija*, Rīga: Biznesa augstskola "Turība".
- Meder, E., & Iske, S. (2009). *Quality assurance by RQCC: how quality is attributed to the relation between learner and e-learning environment*, Proceedings of EDULEARN09 Conference: Barcelona, Spain, July 6-8, 2009.
- Ministru kabineta noteikumi Nr.2, *Noteikumi par valsts akadēmiskās izglītības standartu*. Retrieved January 3, 2002, from: [www.likumi.lv/doc.php?id=57183](http://www.likumi.lv/doc.php?id=57183).
- Ministru kabineta noteikumi Nr. 141, *Noteikumi par valsts pirmā līmeņa profesionālās augstākās izglītības standartu*. Retrieved March 20, 2001, from [www.likumi.lv/doc.php?id=6397](http://www.likumi.lv/doc.php?id=6397)
- Ministru kabineta noteikumi Nr. 481, *Noteikumi par otrā līmeņa profesionālās augstākās izglītības valsts standartu*. Retrieved November 20, 2001, from: [www.likumi.lv/doc.php?id=55887](http://www.likumi.lv/doc.php?id=55887).
- MK noteikumiem Nr.347. *Noteikumi par valsts pirmā līmeņa profesionālās augstākās izglītības standartu*, Retrieved May 29, 2007, from [www.likumi.lv/doc.php?id=6397](http://www.likumi.lv/doc.php?id=6397).
- Rudzinska, I. (2009). Preliminary Evaluation of Process and Result of ESP courses in Latvian HEIs, Language and culture: New Challenges for the teachers of Europe, Selected papers, *Vilnius universiteto leidykla*, pp. 241-251.
- Stiers, J. (2004). Internationalization, intercultural communication and intercultural competence. *Journal of Intercultural Communication*, 11, pp. 1-12. Retrieved from <http://www.immi.se/intercultural>.

