THE USE OF QUALITY MODELS FOR ASSESSING TEST QUALITY IN PROFESSIONAL FOREIGN LANGUAGE COURSE

Dr.Sci.admin. Rudzinska, Ieva
Latvian Academy of Sport Education
e-mail: Ieva.Rudzinska@lspa.lv

“The Need for Educational Research to Champion Freedom, Education and Development for All”, the University of Cádiz, Spain, September 18 - 21, 2012

Introduction

Qualitative assessment is a decisive factor in education in general, and in language learning in particular. The need to pay even more attention to the quality of assessment is caused by increased student competition for budget places in state higher education institutions, and therefore higher demands for fair assessment, as well as by whole society growing awareness about processes in higher education, and therefore higher demand for accountability to wider circle of stakeholders, including society as a whole.

Aims of the research are to select the criteria characterizing professional language test quality, develop model for the evaluation of professional language test quality and evaluate the quality of a professional language Test in the framework of the developed model and selected criteria.

Methods of the research:

- qualitative literature analysis
- quantitative Test score analysis to determine test reliability
- student questionnaire analyses to evaluate Test quality

Theoretical foundations

Language competence is a dynamic combination of professional, communicative and intercultural competences (Luka 2005). Communicative
competence implies successful use of all four language skills with communicative function. Intercultural competence (Stiers 2004, Korhonen 2004) consists of communicative competence together with ability to act in intercultural context, and experience of international work. Professional competence is a combination of communicative and intercultural competences, as well as professional competence, inseparable part of which is professional experience.

Contemporary professional foreign language tests assess the use of skills and competences in situations, resembling target situation as close as possible, besides. Main qualities of professional foreign language tests are validity, reliability and practicality. Valid tests assess what they are intended to, and reliable tests do it in a systematic way.

Latvian higher education laws and regulations indicate that assessment must possess such qualities as summing up positive achievement, reflecting student development; fairness, including openness and clarity of assessment criteria; variety of the forms of assessment and adequacy: test correspondence to the content of the study course, mainly knowledge acquired and skills and competences developed (MK noteikumi Nr. 141 "Noteikumi par valsts pirmā līmeņa profesionālās augstākās izglītības standartu, a.o. Documents).

New development in the sphere of foreign language testing in higher education institutions in EU space of education are GULT task-based tests (GULT Project description, on-line), which use authentic reading and listening materials.

European researchers have developed several models for evaluating the quality in one study course (Lasnier 2007; Meder, Iske 2009; Rudzinska 2009). Quality usually is evaluated in several blocks (objectives, didactic methods, student cognition processes and cooperation, assessment, results, etc.), and according to a list of criteria, among them clarity, adequacy, deep approach, attractiveness, etc. Thus quality model, developed by author (Rudzinska, 2011), consists of 6 blocks and 6 Criteria: Adequacy, Clarity, Attractiveness, Deep approach, Individual work, Cooperation. Integral part of quality models is assessment, which however asks for the inclusion of additional quality criteria. Model, developed by Meder, Iske, for example, includes such additional quality criteria as Objectivity/Subjectivity; Originality/Similarity; orientation to study Process/orientation to study Result (Meder, Iske 2009). Rudzinska
quality model has borrowed the latter criteria from Meder, Iske model (Rudzinska 2009).

Test adequacy means that test tasks reflect the use of language and tasks which test takers would actually perform in target situation (Tratnik 2008). 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, option to choose from several authentic materials.

The inclusion of the criterion of deep approach in quality models emphasizes one of the main aims of higher education: encouraging students to use higher cognitive processes and promote learning for longer term (Dominowski 2002; Biggs 2003). The aims of cognition in the direction from lower to higher are remembering, understanding, applying, analyzing, evaluating and creating (Bloom 1992; Anderson, Kratwohl 2001). Although usually tests test mainly lower level cognitive skills, almost all test tasks can test also higher cognitive processes (Dominowski 2002).

The nature of the last three criteria (Objectivity/Subjectivity; Originality/Similarity; orientation to study Process/orientation to study Result) is a different one, because they combine opposite qualities: qualitative tests should be both original and standard, they should be evaluated both objectively and subjectively, and they should evaluate also process, not only study result.

Quality criteria are comprised of several (2-4) quality indicators.

Test reliability mainly concerns test construct validity. Cronbach alpha values characterize inner consistency of test quality model, and of test quality criteria scales, both should be higher than 0.8.

The correlations between different test tasks should be fairly low: from 0.3 to 0.5, because different Tasks test different aspects of professional foreign language competence. Component (Test task or Item) correlations with the whole test, characterize higher level of order, therefore they should be higher, possibly around 0.7 (Alderson, a.o., 1995). The latter correlations in test theory are called discrimination indexes (D.I.), and are calculated as point-biserial correlations (item-total correlations in SPSS).
Sample of the research:

In the investigation were analyzed 63 Latvian Academy of Sport Pedagogy Year 1 student Tests in Gymnastics, from these 30 students did the questionnaire about Test quality. The sample of 30 students was a convenience sample, which however included most characteristic cases (Geske, Grīnfelds 2006): students from all groups in Year 2, as well as a proportional number of women and men (15).

Student questionnaire

Questionnaire, measuring student opinion about test quality, consisted of 9 questions, 2 quality indicators characterized test clarity (CLA): “Test tasks are clearly formulated”, “Assessment criteria are clearly formulated”; 3: 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 3: test attractiveness (ATT) – “Materials, used in the test, come from authentic problems/situations and 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 in the scale from 1-4, 5th choice was: not appropriate.

Respondents had to evaluate, whether test is objective/subjective, original/similar to others, oriented toward study process/result, choosing from 5 numbers, which extend from one quality to its opposite.

Data analysis

Statistical analysis of the data is performed with SPSS 17.0 software.

Quality criteria comprise several (2-4) quality Indicators. Test quality along the criteria of ADE, CLA and ATT is calculated as median values of the chosen quality Indicators (questions of the questionnaire). 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 help of test task qualitative analysis, which allows implying what cognitive processes have been involved in giving answers to separate Items.

Test result reliability is calculated using Cronbach alpha values and determining Test Criteria, their Indicator and Test Score discrimination indexes D.I.
Results

The results show that the designed model is reliable for test quality evaluation, and give some insights about the quality of the Test.

Reliability of test quality evaluation model

Cronbach’s alpha of the developed Test quality model is 0.88, thus it is higher than the acceptable value. Cronbach’s alpha of separate quality criteria are high enough for ADE and CLA criteria (from 0.73 to 0.76), and not high enough for ATT criterion (0.52), D.I. values are acceptable for ADE and CLA criteria (0.8) and not acceptable for ATT criterion (0.74).

Descriptive statistics for test quality criteria and indicators

Wilcoxon Signed Rank Test has revealed that in student opinion test Clarity, Attractiveness and Adequacy are developed to the same extent. Students have evaluated test quality 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)

The fulfillment of the criterion of Deep approach

Table 1 summarizes cognitive activities the students could have used in performing Task 1 and Tasks 3.
Table 1. Cognitive activities used in performing Test Task 1 and Task 3

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

Table 1 shows that low and medium level cognitive activities are used more than high level ones. To perform Task 1 students have to activate all level cognitive
activities, but in performing Task 3 students are supposed only to remember and apply.

**Descriptive statistics for dichotomous quality criteria**

Median values for dichotomous quality criteria are from 2 to 3. Test is both standard and creative (Figure 2), and learning process and result oriented. However its scoring is more objective than subjective.

![Figure 2. Distribution of student opinion, whether the Test is standard or creative](image)

**Reliability of test scores in Groups A and B**

Cronbach’s alpha for Group A is 0.73 (n=33), for Group B: 0.61 (n=30), D.I. for Task 1, for example, is 0.69 for Group A and 0.58 for group B, similar values are found for other Tasks. Therefore reliability of Version A is higher than that of B.

**Conclusions and discussion**

The reliability of test Version for Group A is higher than that for Group B. To equalize the reliability of both Test Versions some Tasks could be moved from one Test version to another.

The designed Test quality model, consisting of seven quality Criteria can be used as a reliable framework for evaluating test quality. 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 test quality evaluation, indicators, characterizing attractiveness, should be added.

Wilcoxon Signed Rank Test has revealed that in student opinion test Clarity, Attractiveness and Adequacy are developed to the same extent. As regards test compliance with opposite quality criteria, can be concluded that test is almost equally
standard and creative one, and learning process and result oriented. Its scoring however is more objective than subjective. To secure balance between opposite qualities, other control works on the course (presentations, projects, discussions, etc.) should be more subjectively scored.

Deep approach is realized only partly, because low and medium level cognitive activities are used more than high level ones. To secure deeper approach, Task 3 (grammar task) might be incorporated in Tasks 1 and 2.

Another option is to use in the Test authentic reading and listening materials, as practiced in GULT tests. However, the development of such tests can be carried out by joint efforts of staff members, or even higher education institution, testing student foreign language competence in their specialty or professional area.

References


GULT Project description, retrieved 15.05.2010, accessible: *http://gult.ecml.at.*


Ministru kabineta noteikumi Nr. 141 "Noteikumi par valsts pirmā līmeņa profesionālās augstākās izglītības standartu”. 20.03.2001 un to grozījumi, kas izdarīti ar MK noteikumiem Nr.347 „Noteikumi par valsts pirmā līmeņa profesionālās augstākās izglītības standartu” 29.05.2007, accessible: www.likumi.lv/doc.php?id=6397.


P.N.

Personal address: 1 Jūglas iela, apt.70, Rīga, LV – 1024, Latvia, mob.phone: +37129581654, e-mail: Ieva.Rudzinska@lspa.lv;


*This document was added to the Education-line collection on 12 December 2012*