

Running head: UKRAINIAN DEEP

UKRAINIAN DEVELOPMENT OF AN ENGLISH EXAM PROJECT

UNDERWAY

Igor Romanyshyn

Vasyl Stefanyk Precarpathian National University (Ukraine)

Abstract

In this paper I refer to the background of the Ukrainian Development of an English Exam Project (Ukrainian DEEP) aimed at training specialists in the field of “independent” foreign language testing. I show what we, participants, learned and did during the three rounds of training. The British Council and the Ministry of Education and Science of Ukraine jointly support this project. It is of great importance for improving the national testing system. The latter is possible through raising the project participants’ awareness of principles and practice of language testing and through involving Ukrainian stakeholders in the project.

I Introduction

Testing is too important to be left to testers

Charles Alderson

The project started in February 2006. There were over 30 participants from different parts of Ukraine. At that time we all thought we knew much about language testing as most of us were professional teachers working in secondary schools, technical schools or institutions of higher learning. We often used tests to assess our students' progress in or proficiency of the English language. Some of us even designed tests for admission purposes for Examination Boards of our schools. What we expected from the project – this is my subjective opinion driven largely by the then policy of the Ministry of Education – was that we would learn how to organize and administer “independent” nation-wide school leaving exams in English. At that time we felt confident that our knowledge and experience as university professors and scholars or high school teachers were sufficient to contribute much to the project. That was on the first day of the seminar when many of us realized we knew little about testing.

II What we learned and did

The first round of seminars was held in Kyiv on 13 –17 February 2006. It was an intensive introductory course in Language Testing and Common European Framework of Reference (CEF). Charles Boyle, Zoltan Toszegi and

Patricia Aresvik, our trainers from Delaso (Defence Language Solutions UK Ltd), introduced us to

- The Common Reference of Languages, its levels and rating scales
- Core concepts and language testing terminology such as reliability, validity, washback, the language testing cycle, intra-rater reliability, etc.
- Test specifications
- Task format and item types for testing listening, speaking, reading and writing.

Besides acquiring theoretical knowledge through the trainers' presentations and readings from Alderson (Alderson, Figueras, Kuijper, Nold, Takala, & Tardieu, 2006), Fulcher (2004), North (2004) and CEF (2001), we had practice of group and "syndicates" work on developing test specifications, analysing tasks, rating students' writing papers and designing our own test items. We also carried out a task on construct retrieval using the DIALANG Self-assessment statements related to Reading (the *Common European Framework of Reference Manual*, 2001, p. 231).

For homework we were assigned to read Section 4.4.2.2 of the CEF Manual (ibid, pp. 68-71) and, using the five tables in that section, to retrieve a comprehensive construct of Reading that was to be inserted in the provided Microsoft Excel spreadsheet. We were to familiarise ourselves with a number of issues such as: text types, domains and purposes for reading, skills and subskills,

scales, levels and descriptors of reading competence and other topics. We all worked hard to complete the task.

During the second round of seminars that was held in Kyiv on 26 April-6 May 2006, the project participants read from Alderson and Cseresznyes (n.d.), Nikolov (1999), Peirce (1994), Weir (2005) and Figueras (Figueras, North, Takala, Verhelst, & Avermaet, 2005). On the one hand, it took all the evenings to prepare for reading tutorials. On the other hand, we acquired invaluable knowledge of how to select texts and develop items for reading comprehension (e.g. the principle of “authentic language” of the text and items, the “defensibility of items”, the pre-testing of items, their revision, statistical analysis of the results of the item performing, etc. as found in Peirce (1994)). From M. Nikolov’s (1999) *Classroom Observation Project* we got to know that to develop a new school-leaving examination it is necessary first “to gain realistic insights into average teaching conditions, pedagogical processes, and task types teachers use in secondary schools”, especially on the peripheries, so as to make conclusions and recommendations both for designers of the new test (e.g. how to reflect the syllabus in the test and/or to link the latter to the CEF) and teachers (how to improve the teaching/learning process and prepare students for the ordeal).

Though the CEF is increasingly becoming the standard for language curricula, textbooks and other teaching materials, as well as for modern European language examinations, assessment procedures and courses (North,

2004; Alderson and Cseresznyes), some authors (Weir, 2005; Alderson et al, 2006) state that there are limitations of the CEF for developing comparable examinations and tests. They found inconsistencies in the CEF scales and grids, terminology, purposes for completing test tasks, response format, time constraints, channel for communication, discourse modes, text length, topic, lexical and structural competences and other aspects that make up the construct of a test. Our trainers split us into small groups to carry on the home task on reading construct development. As it turned out, each participant had a different understanding of the reading components. As a result, it took us a number of drafts and a few days to complete only Levels A1, A2, B1 and B2. Alongside we were involved in test specifications development, but it did not take us far either. There were also organized workshops in which the participants practiced text and task selection. Divided into six work groups, we selected 36 authentic texts for Levels B1 and B2 and began designing tasks for them. Each group had to do six tasks. All the tasks, including rubrics, estimated completion time and marking key, had to be in final draft format before the end of the seminar. Moreover, they were to be revised and improved by other participants. Several of us, together with Olena Gorshenova and Olexandr Shalenko (representatives of the British Council Ukraine) were also engaged in designing a questionnaire for determining reading interests of secondary pupils. The final draft of the questionnaire came to include such aspects as: place and conditions under which

a pupil reads in English, types of texts, literature genres, time spent both on reading in a mother tongue (Ukrainian and/or Russian) and in English, number of pages per week a pupil reads both in a mother tongue and in English and a self-assessment grid in reading.

Thus, for homework we were given two major tasks:

- To conduct the questioning of school-leavers about their reading habits and interests
- To pilot our reading texts and tasks on the same group of pupils.

Both tasks required much and intensive work on the part of the project participants, especially the second one – the completing of the four stages of the piloting: (1) administering the piloting in secondary schools of various types under required conditions (e.g. each task had to be pre-tested on at least 10 school pupils in their final year(s) who were believed to be at CEF levels B1 or B2; there were to be copies of the task/answer sheets with instructions for each student, the task administrator had to make notes of any issues that arose on the task evaluation sheet and to carry out orally or in writing the pupils' feedback after they had finished the task); (2) collating the pupils' responses (this required marking the responses as per the Answer Key and recording the results on the Task Analysis grid); (3) analysing the results (for this we had to analyse the data on our Task Analysis grid, consider the learners' feedback comments, consider how the task could be modified as to the problems that had arisen, to note our ideas and recommendations for modifying the different components of the task

on the Task Evaluation sheet, etc.), and (4) modifying the tasks or rejecting them. Each member of a group had to do the first three tasks individually, while the last one was to be done after the group analysis of the Task Analysis grids, Task Evaluation sheets and input from students' feedback. I have to admit that all the groups managed to complete the tasks and get prepared for the next presentation mostly due to understanding of the importance of the new English exam development project and the assistance of schoolteachers and school administration (though there were a few cases when the latter did not allow the conducting of the questioning and/or the piloting because there was no letter of support from the Ministry). Here I can't but agree with Charles Alderson's words: "Testing is too important to be left to testers" mentioned in his plenary address to the Third Annual Conference on Current Trends in English Language Testing, United Arab Emirates University, Al Ain and Zayed University, Dubai Campus (personal communication, May, 1999).

The third round of seminars took place in Morshyn, Lviv region, on 26 July-2 August 2006. As during the previous seminars, we had reading tutorials based mainly on the articles of Cheng and Qi (2006), Qi (2005) and Yamashita (2003) concerning the nature of reading, reading skills and strategies, task formats, washback, etc. Besides, we were involved in final group preparations followed by presentations of our tasks. In the presentations we stressed some drawbacks that the piloting had revealed: the format of some rubrics appeared to be long and confusing for the pupils; some items were badly

worded, thus wrongly understood by the pupils; some questions overlapped, were too general or too easy; there were acceptable though unpredicted responses, etc. We took all of them into consideration and suggested major or minor improvements for the task or its rejection.

I believe all of us benefited much from the presentations as well as from getting practical advice on designing and conducting the piloting of a new questionnaire, doing the think-aloud protocols and constructing multiple-choice items.

III What we should do

On the final day of the seminar we were asked to decide what area of testing we would like to choose. I, for example, answered writing specifications and items, doing the piloting, designing questionnaires and promoting the project. These and other fields of testing, as well as the other language competences – speaking, listening and writing – need to become the focus of the further rounds of seminars on Principles and Practice of Language Testing. We look forward to participating in them.

IV Conclusion

If the Ukrainian Ministry of Education wants to produce and administer a quality language examination at the national level, it has to invest money in it. It has to involve in the project language testing specialists, teachers, policy

makers, resource providers and other stakeholders. It has to closely cooperate with the British Council Ukraine in both introducing a tester-training program within the framework of refresher courses for foreign language teachers and promoting the project, in particular through providing its participants with necessary assistance.

Y Acknowledgements

On behalf of those involved in the project I would like to acknowledge all the individuals and institutions that have helped us to participate in it. We are grateful to international EFL methodologists whose ideas and activities contributed to the project. We would like to thank our trainers whose presentations and tutorials made us all aware of language testing, whose reviews helped in developing the tasks and reading questionnaires. We would like to thank administrators and students of the schools that were the testing ground for the piloting and, of course, our families for allowing us the possibility to stay in the project.

YI References

- Alderson, J. & Cseresnyes, M. (n.d.) *Into Europe. Prepare for modern English exams. Reading and use of English*. Budapest: Teleki Laszlo Foundation.
- Alderson, J., Figueras, N., Kuijper, H., Nold, G., Takala, S. & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR Construct Project. *Language Assessment Quarterly*, 3: 3 – 30.
- Cheng, L. & Qi, L. (2006). Description and examination of the national matriculation English test. *Language Assessment Quarterly*, 3, 53-70.
- Council of Europe. (2001). *Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press.
- Figueras, N., North, B., Takala, S., Verhelst, N., & Avermaet, P. (2005). Relating examinations to the Common European Framework: a manual. *Language Testing*, 22, 261 – 279.
- Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, 1, 253-256.
- Nikolov, M. (1999). Classroom observation project. In *English Language Education in Hungary* (pp. 221 - 245). British Council Hungary.
- North, B. (2004). Relating assessments, examinations and courses to the CEF. In K. Morrow (Ed.) *Insights from the Common European Framework* (pp. 77 – 90). Oxford University Press.
- Peirce, N. (1994). The Test of English as a Foreign Language: developing items for reading comprehension. In *From Testing to Assessment* (pp. 39 – 60). Longman.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22, 142 – 173.
- Yamashita, J. (2003). Process of taking a gap-filling test: comparison of skilled and less skilled EFL readers. *Language Testing*, 20, 267-293.
- Weir, C. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing* 22, 281 – 300.