

Sonya Toncheva, Daniela Zlateva, Martin Ziarati  
Nikola Vaptsarov Naval Academy  
Bulgaria

## The importance of developing test specification in the process of the Enhanced Oral Test design

### Abstract

*The paper will explain the factors taken into consideration in the development of the Enhanced Oral Test, which is one of the outcomes of the EU Funded Leonardo MarTEL PLUS project. The development of an International test of the speaking skills of seafarers will be by its nature a very high stakes test and should be based on solid testing principles to ensure validity and reliability of the test scores. Therefore, the test specifications are important in the establishment of the test's construct validity. They are also important for test writers in developing new versions of the test to ensure its sustainability.*

**Keywords:** test specifications, Enhanced Oral Test, construct validity, sustainability, test design, MarTEL, Maritime English test

### 1. Introduction

Developing a speaking test which will be used for job selection purposes is a very demanding and highly responsible activity considering the fact that important decisions will be made based on the test results. However, as experts in the field of testing agree, before any test is put into practice, its quality and sustainability should be examined carefully to provide evidence that the test can be used as a valid and reliable measurement instrument. Test developers should provide comprehensive answers to a number of universal questions related to all stages of the test design process. These answers should be reflected in the test specifications document which will guide the entire process of test development from specific tasks to complete tests to ensure a balance between different aspects of test usefulness (reliability, construct validity, authenticity, interactiveness, impact, practicality) and find the most acceptable solution in the specific context.

The paper will identify the potential users of the test specifications and specify the particular testing context. Then, the most important questions at each stage of the test design process will be addressed. Finally, conclusions will be drawn based on what researchers suggest and our own experience.

### 2. Users of Test Specifications

A number of researchers in the field of testing have contributed to the structure and purpose of the test specifications (e.g. Lynch and Davidson, 1994; Alderson *et al.*, 1995, Bachman and Palmer, 1996). They view the structure of this document from different standpoints; however they agree that the different versions should be aimed at different audience. As the aim of this paper is to focus on the importance of

this document in the test design process, the different versions will not be discussed. Nevertheless, targeted users will be identified.

First, test developers will need the detailed version of the test specifications to use it as a guide for the writing of new tests and ensure sustainability. Test items moderators will also need to consult the document when they review the work done by the test developers. When test validity is established the evaluators will need to refer to the document, too.

Another group of test specifications users are the test-takers themselves who want to become familiar with the test structure, task types, expected performance, time allotment, assessment criteria. Teachers might also be interested in the same issues in order to prepare their students for the test or include similar content in their lessons.

Furthermore, information for public use may be needed by university admission officers, company managers who will have to select a valid test for their needs.

### **3. The testing context**

Assessing linguistic competence in Maritime English adequately and reliably at internationally recognized levels has been set forth in recent years as a major issue because it reaches out equally to merchant marine officers, cadets and students, as well as Maritime English Training (MET) institutions, maritime administrations, ship owners, etc. Indeed, all the above mentioned parties have come to recognise the need of developing exam systems evaluating spoken competence (Logie, 2011). and conducting Maritime English oral tests to this effect. Furthermore, the necessity to ensure effective communication (in both written and oral form) in its diverse manifestations in various nautical and technical spheres has been explicitly expressed in the Manila amendments (2010) to the STCW Convention 1978/95 STCW (2011).

Based on feedback received from different parties and in response to the need of developing a more comprehensive process for the evaluation of oral competence, as raised in the 2010 IMO STW 41 meeting, the MarTEL Plus project set, as one of its goals, to embark on enhancing the speaking part of the MarTEL test of Maritime English language proficiency. The MarTEL projects were developed under the EU Leonardo da Vinci funding stream, as part of the EU's Lifelong Learning Programme (MarTEL, 2007). They envisaged this as a complement to the existing MarTEL standards, with a two-tier system, including the current MarTEL speaking section in the Phase tests, plus a separate one to one oral examination – the Enhanced Oral Test (EOT) (MarTEL Plus, 2010).

Testing the speaking skills of seafarers is by its nature of very high stakes as it will affect a large number of people. Therefore, the approach to test design was determined by the basic principles of test development. Writing the guiding test document – the EOT test specifications implied asking and answering all relevant questions at each stage of the process.

### **4. Design stage**

This is the first stage of test development which involves gathering important information to write the test specifications. The most important aspects to consider are the test purpose and the definition of the construct. Test developers should also identify the target population, i.e. who the test-takers are and their real world specific speaking needs.

#### **4.1 Purpose of the test**

Identifying the test purpose, i.e. how the test should be used and who it is intended for is related to the validity of test results. Validity is a complex phenomenon and has several aspects. One of them relates to the correctness of the inferences or decisions made on the basis of test results. Each test can be valid only for the purpose it was designed for. It means that as a proficiency test the EOT should not be used as a diagnostic or progress test, for example. The test developers have made it clear in the Test Specifications how the test will be used. All attempts to use the EOT for other purposes should be referred to as inappropriate or test misuse.

#### **4.2 Defining the construct**

The construct is a term which refers to the definition of what the test is supposed to measure. The test specifications should clearly state the test developers' definition of what speaking means for seafarers and what aspects of it the test will attempt to assess. To put it simply, the test is valid when it measures what it is intended to measure. Therefore, research on the specific oral capabilities or skills required by the maritime industry was carried out to identify the target language use domain of the EOT as the test tasks must represent the language skills needed in this specific domain. Following the Common European Framework (CEFR) (CEFR, 2011), five levels of speaking proficiency were developed and called Martel Plus Level Descriptors. These descriptors broadly define the specific purpose speaking skills needed by the target test population.

As the Model course (3.17) on Maritime English presents the IMO requirements on use of Maritime English in professional context a list of topics was created to provide the job-related context of the speaking tasks.

Tests, in general, are only samples from a content domain. One would not expect a test of specific speaking ability to measure every single aspect or function of speaking involved in a particular context. The test developers should make sure that the test is as representative of the content domain as possible. However, if one or more important aspects of the content domain are not included in the test, then construct validity is under question and this is known as "construct underrepresentation". An example of this would be to omit from the definition of the construct the ability of seafarers to communicate using the IMO's Standard Marine Communication Phrases (SMCP). The test will not be valid for its intended use. In addition, the inferences based on the test results will be far from valid. When safety at sea is at stake, this is a real danger.

Another potential danger to construct validity would be what test evaluators call "construct-irrelevance". This is observed when a test task assesses something else than it is supposed to test. For example, in a role play where the test-taker assumes the role of the Officer of the Watch on board a ship, the examiner gives instructions about the conversation in such a complex language using low frequency vocabulary and grammar patterns that the test-taker fails to understand the instructions and to complete the task. What is assessed in this case is not the test-taker's ability to demonstrate language skills to perform a work-related task but whether s/he can understand the examiner's highly complex language.

### **5. Operationalization stage**

At this stage the test developers use the information from the design stage to create guidance for the development of specific tasks and complete tests. The most important questions to ask and answer during

this process are how to make test-takers demonstrate the specific speaking abilities defined in the construct, what criteria will be used for assessing these abilities and how the test score will be formed.

## **5.1 Developing task specifications**

### **5.1.1 Major concerns on the EOT speaking tasks selection**

Research findings show that it is difficult to find suitable and novel tasks that test communicative ability alone and not intellectual capacity, educational and general knowledge or maturity and experience of life.

In addition, the choice of the type of assessment is limited to construct based and task based where the latter is especially used in professional contexts as the scores give information about the test-taker's ability to deal with the demands of the situation. Researchers do not look at the two perspectives as 'conflicting' (Luoma, 2010). Therefore, combining elements of the two appeared to be the tool that satisfied the needs of the EOT maritime context.

Another issue to consider refers to ethics. Being fair to all test-takers is a major matter of concern for all test developers and examination boards. This is the reason why some formats come with the accompanying test materials, e.g. sample materials, preparation materials, etc. to provide conditions for fair testing. Therefore, ample Tasks were developed, piloted and incorporated in the EOT Task Specifications document. This would give equal opportunities to all those interested in taking the test to become familiar with the task types, expected response and assessment criteria.

Last but not least, the washback effect (or backwash as used in the general education field) should not be ignored. The notion of 'washback' refers to the influence that tests have on teaching and learning. Different aspects of influence have been discussed in different educational settings at different times in history due to the fact that testing is not an isolated event (Shohamy, 1993).

Furthermore, researchers suggest that 'high-stakes tests' would have more impact than low-stakes tests (Alderson and Wall, 1993). If we consider the new EOT a high-stakes test, we should then be aware of factors such as the status of the subject, i.e. English within the curriculum, the nature of teaching materials, teacher experience and teacher training, teacher awareness of the nature of the test as they all would affect the amount and type of washback. New tests do not necessarily influence the curriculum in a positive way as changes do not happen overnight and teachers do not always feel ready to implement changes. In his study on the washback effect of the Revised Use of English Test, Lam concludes that it is not sufficient to change exams: "The challenge is to change the teaching culture, to open teachers' eyes to the possibilities of exploiting the exam to achieve positive and worthwhile educational goals" (Lam, 1994).

One conclusion based on washback research findings is that there is a complex interaction between tests on one hand and language teachers, material writers and syllabus designers on the other hand and we should be aware of this.

### **5.1.2 Major aspects of test task development**

There are seven task types employed in the EOT. To ensure sustainability test developers must provide information on how the new versions of the test will be developed. A set of task characteristics for each individual task guides test writers and includes the following:

- the definition of the construct to be assessed (the speaking skills),
- task difficulty,

- genre,
- stimulus material,
- the setting of the task,
- time allotment,
- type of interaction involved,
- instructions for responding to the task,
- type of input (specifying the quality of visual prompts),
- assessment criteria

Without this information which will serve as a template for task design test developers will find it difficult to be consistent in providing comprehensive guidelines for writing new test versions and being fair to all test-takers.

## **5.2 Specifying the assessment criteria**

The next set of questions which the test developers must answer relates to the criteria for correctness. They should provide information about how the criteria were developed and how the rating scale is used to measure the construct. This information is an important part of the test specifications as it will be used for validating the rating scales. The rating scale (analytic or holistic) and the assessment criteria should be used during examiner training sessions to ensure reliability of marking.

In assessing speaking skills in general, reliability is a big problem area. The test specifications may have the assessment criteria but it is important to ensure that the same rater or assessor will apply the scales in the same way on different days or at different times of the day (intra-rater reliability) and the different assessors apply the scales in the same way (inter-rater reliability). These issues can be addressed by assessor training workshops to minimize the assessor variables (background, experience, expectations, etc.) which can be very influential in determining the scores.

The EOT developers have produced a separate document accompanying the test specifications. This document will serve as guidelines for assessor training with the purpose of achieving consistency of measurement. During the training process part of the test specifications may undergo some changes in the wording of a certain criteria, for example, so that assessors reach agreement on the interpretation of some or all assessment criteria.

## **6. Administration Stage**

At this stage pre-testing is carried out to collect information and feedback about the test. This information is then analysed and discussed by test developers. The feedback received may make it necessary to return to a previous stage to rectify a problem. This in turn, may lead to making changes and reviewing the test specifications.

## **7. Conclusion**

Developing speaking tests for maritime purposes should be done with greatest care possible and by a team of test developers including a subject matter specialist, somebody with a testing background and a statistician.

Test writers should follow all stages of test design and produce the accompanying test specifications to provide a system for test development.

Time and efforts should not be sacrificed to review the test and the entire test documentation after pre-testing so that the test becomes a valid and reliable measurement instrument.

We owe this to seafarers.

### References

1. Alderson, J.C. and Wall, D. (1993) 'Does Washback Exist?' *Applied Linguistics*, 14(2): 115-29
2. Alderson, J.C., Clapham, C. and Wall, D. (1995) *Language Test Construction and Evaluation*. Cambridge: CUP
3. Bachman, L. F. and Palmer, A. (1996) *Language Testing in Practice*. Oxford: OUP
4. Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR), CUP, 2001, pp 58-59.
5. Lam, H.P. (1994) Methodology washback – an insider's view. In D.Nunan, R.Berry, & V.Berry (Eds.), *Bringing about change in Language Education: Proceedings of the International Language in Education Conference 1994* (83-102). Hong Kong: University of Hong Kong
6. Luoma, S. (2004) *Assessing Speaking*, Cambridge Language assessment Series, Cambridge University Press, p.42
7. Logie., (2011). Whose culture? The impact of language and culture on safety and compliance at sea Retrieved February 4-th 2011 from <http://www.ukpandi.com/fileadmin/uploads/uk-pi/LP%20Documents/Industry%20Reports/Alert/Alert14.pdf>.
8. Lynch, B. and Davidson, F. (1994). Criterion-referenced language test development: linking curricula, teachers, and tests. *TESOL Quarterly* 28(4)
9. MarTEL Project 'Maritime Test of English Language' (2007-9), EU Leonardo Project, UK/07/LLP-LdV/TOI-049, 2007-2009.
10. MarTEL Plus Project 'Maritime Test of English Language Plus' (2010-12), EU Leonardo Project, UK/010/LLP-LdV/TOI-342, 2010-2012
11. Model Course 3.17 Maritime English (2009), IMO Publication, CPI Books Limited, UK, second edition, ISBN 978-92-801-1502-4
12. Shohamy, E. (2009), *The power of tests: The impact of language tests on teaching and learning*. NFLC *Occasional Paper*. Washington, DC: National Foreign Language Center,
13. Subcommittee on Standards of Training and Watchkeeping (2011), Report to the Maritime Safety Committee Retrieved February 24th 2011 from <http://www.uscg.mil/imo/stw/docs/stw41-report.pdf>.

### Author's Bio-Note

#### **Assoc. Prof. PhD Sonya Toncheva, MSc English Philology PhD Maritime Terminology**

Head of Language Department, Nikola Vaptsarov Naval Academy Varna

Email: [info@naval-acad.bg](mailto:info@naval-acad.bg) Tel: 052 632 015, Fax: 052/303 163

Address: 73 V.Drumev Str., 9026 Varna

**Assoc. Prof. PhD Sonya Toncheva** is Head of the English Language Department at the Naval Academy, Varna. She has been teaching cadets and students of different specialties – Navigation, Port Management and Operation and Ship Engineers. She has taught English for SAR operations in the International University in

Triest and for the London Institute of Ship Brokers. Her PhD thesis is devoted to a comparison between the English and Bulgarian Maritime Terminology. She has undergone specializations in UK, Slovenia and Ukraine. She has specialised in USA in a course on Management of English Teaching. Sonya Toncheva is the author of books and a dictionary for teaching Maritime English and has participated in many IMEC, GAME and IMLA conferences of which she is a member. About her work she has been awarded a »Lady Cross« and in 2009 the Sign of the town of Varna for education.

**Daniela Zlateva, MSc English Philology, MA Language Testing**

Senior Lecturer, Nikola Vaptsarov Naval Academy Varna  
Email: [info@naval-acad.bg](mailto:info@naval-acad.bg) Tel: 052 632 015, Fax: 052/552 225  
Address: 73 V.Drumev Str., 9026 Varna

**Daniela Zlateva** is a senior lecturer in English at the Department of Post-Graduate Training at the Naval Academy, Varna. Her teaching experience involves participation in a number of projects for teaching and testing naval officers, providing consultancy work for the Ministry of Defence, developing materials for self-study and distance learning. She has an MA in Language Testing from Lancaster University, UK and has specialised in the US in test development and methodology of teaching.

**Dr Martin Ziarati, BA Hons (Business Economics), PhD (Engineering)**

Director, Centre for Factories of the Future  
Email: [martin.ziarati@c4ff.co.uk](mailto:martin.ziarati@c4ff.co.uk) Tel: +44 (0) 2476236734, Fax: +44 (0) 2476470060  
Address: Centre for Factories of the Future, Barclays Venture Centre, Sir William Lyons Road, Coventry CV4 7EZ, United Kingdom

**Dr Martin Ziarati** is the Director and Head of maritime education and training at the Centre for Factories of the Future (C4FF).

He is the coordinator of the MarEdu and MariFuture networks. He has undertaken coordination activities for a number of EU funded projects. He is presently the project director for six ongoing EU funded projects including two Maritime English focused projects: MarTEL Plus and CAPTAINS. C4FF lead the MarTEL Plus project which intends to establish international standards for Maritime English with a total of 22 EU partners.

He has written a number of international refereed papers in the area of Maritime English. He has written a number of articles printed in international maritime publications in the area of Maritime Communications. He is a member of the Excellence Club, represented by leading innovative companies in the region and a personal member of the EU Research and Development funding group, both established by the regional development agency.

He is a Working Group Member, MILC Technology & Innovation Group, MILC (Marine Industries Leadership Council (UK's forum for the industries' key stakeholders supporting UK Government Policy Making). He is a visiting Senior Research Fellow and advisor on a number of ongoing PhDs at De Montfort University, UK. His recent activities include being a member of the organisation committee for an International workshop and session chairs for a number of international conferences.