

Guidelines for the Development of Language for Specific Purposes Tests

A supplement to the Manual for Language Test Development and Examining

Produced by ALTE



© ALTE, 2018

All correspondence concerning this publication or the reproduction or translation of all or part of the document should be addressed to the ALTE Secretariat (secretariat@ALTE.org).

Contents

Foreword	2		
1 Fundamental considerations	3		
1.1 Scope	3		
1.2 Audience and usability.....	3		
1.3 Defining and classifying LSP tests.....	3		
1.3.1 Issues in defining and classifying LSP tests.....	3		
1.3.2 Possible ways to define and classify LSP tests.....	3		
1.4 Structure of the guidelines	4		
2 Developing the LSP test	5		
2.1 The process of developing the LSP test	5		
2.2 The decision to provide a test	5		
2.3 Planning	5		
2.3.1 Assessment experts	5		
2.3.2 Subject and context experts	5		
2.3.3 Roles of the subject and context experts	5		
2.3.4 A special group of context experts: LSP teachers	6		
2.3.5 Needs analysis.....	7		
2.4 Design	8		
2.4.1 Initial considerations.....	8		
2.4.2 How to balance requirements with practical considerations	9		
2.4.3 Test specifications.....	9		
2.5 Try-out	9		
2.6 Informing stakeholders.....	9		
2.7 Key questions.....	10		
3 Assembling LSP tests	11		
3.1 The process of assembling tests	11		
3.2 Preliminary steps.....	11		
3.2.1 Item writer recruitment and training	11		
3.2.2 Managing materials.....	12		
3.3 Producing materials.....	12		
3.3.1 Assessing requirements.....	12		
3.3.2 Commissioning	12		
3.4 Quality control.....	13		
3.4.1 Editing new materials	13		
3.4.2 Piloting, pretesting and trialling	13		
3.4.3 Review of items	13		
3.5 Constructing tests	13		
3.6 Key questions.....	13		
4 Delivering tests	15		
4.1 Aims of delivering tests.....	15		
4.2 The process of delivering tests	15		
4.2.1 Arranging venues	15		
4.2.2 Registering test takers	15		
4.2.3 Sending materials	15		
4.2.4 Administering the test	15		
4.2.5 Returning materials	15		
4.3 Key questions.....	16		
5 Marking, grading and reporting of results	17		
5.1 Marking	17		
5.1.1 Clerical marking	17		
5.1.2 Machine marking.....	17		
5.1.3 Rating.....	17		
5.1.4 Rating criteria	18		
5.2 Grading.....	19		
5.3 Reporting of results	19		
5.4 Key questions.....	19		
6 Monitoring and review	20		
6.1 Routine monitoring.....	20		
6.2 Periodic test review	20		
6.3 What to look at in monitoring and review.....	20		
6.4 Key questions.....	21		
References and further reading	22		
Acknowledgements	23		

Foreword

I am pleased to have the opportunity to write the Foreword to this supplement to the Manual for Language Test Development and Examining for use with the CEFR (2011). The supplement is a timely addition to the Manual and evidence of the continuing role that ALTE plays as an INGO with participatory status within the Council of Europe.

ALTE has a long-standing collaborative relationship with the Council's Language Policy Division dating back to the early 1990s. ALTE and individual ALTE Members have contributed to the development and evolution of the Common European Framework of Reference for Languages (CEFR) since the original authoring group and consultative committees were established in 1991/2.

Following the release of the draft version of the CEFR in 1996, ALTE was invited to produce the Guidelines for Test Development and Examining to accompany the Framework. This work was coordinated by Dr Michael Milanovic, who was the Manager of ALTE at that time, and the first set of ALTE/Council of Europe guidelines was published in 1997.

After the finalised version of the CEFR appeared in 2001, the Language Policy Division encouraged the development of additional resources to support appropriate uses of the CEFR in different contexts of learning, teaching and assessment. These resources have become known as the CEFR 'toolkit', which now includes a range of 'tools' to explain, exemplify, support and enable appropriate uses of the Framework. Examples of these resources include: a Relating Language Examinations to the Common European Framework of Reference for Languages (CEFR): Learning, Teaching, Assessment: A Manual with its Reference Supplement; Illustrative Descriptors and Calibrated Sample Materials; Content Checklists and Grids; and the Reference Level Descriptions (RLDs) – see the Council of Europe website and reference to Profile Deutsch, English Profile, etc.

ALTE has participated in many of these projects and continues to engage with the Council of Europe through ALTE's Special Interest Groups (SIGs). In 2008, ALTE was invited to revise the first guidelines for test development and examining to produce an updated Manual taking into account developments in the field of assessment and cross-referencing to the CEFR and its toolkit more effectively. This work, coordinated over a period of 24 months by the ALTE Secretariat and the ALTE CEFR SIG, involved consultation with nearly 100 participants from across the ALTE membership and beyond. The introduction to the Manual by Milanovic outlines the main aims and provides practical guidance to readers on how to use it effectively in their own assessment contexts.

Hard on the heels of the publication of that Manual, ALTE's Language for Specific Purposes (LSP) SIG started discussing the possibility of producing supplementary materials to enhance its use for providers of language for specific purposes (LSP) examinations. While the LSP group members acknowledged that the Manual in conjunction with the CEFR provides useful guidance for developing tests of general language proficiency, when it comes to LSP tests and especially 'strong (narrow) LSP tests', they acknowledged that these can be more difficult to develop without further guidance that makes reference to specific domains of use.

Rather than starting from scratch, the LSP SIG decided to build on the strengths of the Manual itself, one of which is that it very clearly presents the stages of test development and administration which apply to any kind of language test. That is why the supplement is still organised around the same core guidelines and process-oriented cycles.

Where the supplement comes into its own is with regard to domain specificity. The additional guidelines highlight this kind of specificity throughout all stages of the process by emphasising a number of essential concerns. These concerns include the vital importance of involving subject matter and context experts, the increased need to achieve situational and interactional authenticity and other considerations required in adapting the assessment system to the target language use (TLU) contexts. The aim is, therefore, to combine the common principles of language assessment with specific ways of complying with the requirements of domain-specific stakeholders.

As with all ALTE projects, this work has been undertaken with a clear focus on multilingualism, both personal (plurilingualism) and in relation to policy-making at institutional and societal levels. A cursory glance at the literature on LSP confirms that most of the research has been published in English and with reference to English language uses for specific purposes (ESP). The material in the supplement is therefore a welcome addition to the literature as it takes into consideration the diverse, multilingual contexts of assessment that ALTE represents and is designed to be applicable to all languages. Indeed, the examples used in the guidelines actually refer to different testing contexts in terms of the target language being learned and assessed. In many cases, the main reasons for learning a less widely used language are related to very specific purposes linked to work, study or social mobility.

In keeping with the CEFR concept and the evolving practices associated with its uses, I fully expect these guidelines to be extended and adapted according to stakeholder needs and as they are put to good use in 'real-life' testing programmes. I look forward, therefore, to new developments in future and to further additions to the CEFR toolkit produced by ALTE members.

Nick Saville, PhD
ALTE Secretary-General
December 2017

1 Fundamental considerations

1.1 Scope

Guidelines for the Development of Language for Specific Purposes Tests is a document designed to accompany the Manual for Language Test Development and Examining produced by the CEFR Special Interest Group (SIG) of ALTE for the Council of Europe. The purpose of these guidelines is to highlight issues specific to language for specific purposes (LSP) tests, while adhering to the basic test development principles/procedures presented in the Manual. The guidelines were developed by the LSP SIG of ALTE.

1.2 Audience and usability

The guidelines are primarily aimed at test providers interested in developing LSP tests. The test developer is advised to follow the steps in the Manual (Chapters 2 to 6) for the development and administration of general language tests and consult, at the same time, the specific LSP test information.

1.3 Defining and classifying LSP tests

LSP tests derive both their test content and test method from an analysis of the target language use (TLU) domain that is important to the test takers. The purpose of the TLU analysis is to predict language performance in specific contexts, such as a workplace (Douglas, 2001).

1.3.1 Issues in defining and classifying LSP tests

Producing a clear definition of what constitutes an LSP test versus a general purpose language test is not straightforward. Specialists regard them as a continuum of specificity: from very general to very specific tests (Douglas, 2000; O'Sullivan, 2006).

Some of the most important problems we need to address in defining LSP tests are:

- The variable extent to which mirroring the authentic language, tasks and contexts of the TLU domain is necessary. This makes it sometimes difficult to distinguish between an LSP test and more general performance-based testing.
- The crossover between language and content. For example, to what extent is a 'German for Business' exam testing candidates' knowledge of German, and to what extent is their knowledge of business being tested?
- The extent to which the involvement of specialists/informants in the domain is necessary for the development, construction, rating and grading of LSP tests – aspects which impact the validity of the test.

1.3.2 Possible ways to define and classify LSP tests

LSP tests can be classified by their purpose or by their content:

- by their purpose, for example: tests for academic purposes, tests for work purposes, tests for 'legal acts' (migration, citizenship), literacy tests;
- by their content, for example: broad and narrow LSP tests.

Broad LSP tests

A broad definition would include LSP tests situated closer to the general end of the continuum (O'Sullivan, 2006), for example tests related to Language for Business, or Language for Academic Purposes. For these more general LSP tests, it might not be necessary to include subject and context experts in test development, test construction, rating and review (although consultation with stakeholders and end users of test results is still very important).

As an example, see Occupational English Test (OET) (Australia), designed for a wide range of healthcare professions: www.occupationalenglishtest.org

For sample test materials of the OET:

www.occupationalenglishtest.org/Display.aspx?tabid=2425

For another example, see Zakelijk Professioneel (PROF) (Belgium and the Netherlands), a B2 level exam for those who need Dutch in an occupational context, more specifically in healthcare and administration: cnavt.org/en/introduction

Narrow LSP tests

A narrower definition would restrict LSP tests to those focusing on a specific domain, for example Law, Finance, Medicine, and Aviation. For narrow LSP tests it would be important to involve subject and context experts at each stage of the development and construction process, and possibly also in rating. However, we should not forget that these LSP tests are still language tests and that language testing professionals would still keep the lead in the development and construction process.

As an example, see Canadian English Language Benchmark Assessment for Nurses (CELBAN):
www.celbancentre.ca

These guidelines focus predominantly on narrow LSP tests, as they are more distinct from general language tests, and present more challenges to the test developer. However, broader LSP tests are also considered and referred to throughout the guidelines where appropriate.

1.4 Structure of the guidelines

The Guidelines for the Development of Language for Specific Purposes Tests mirror the structure of the Manual for Language Test Development and Examining and hence contain the following sections: Developing the LSP test; Assembling LSP tests; Delivering tests; Marking, grading and reporting of results; Monitoring and review. The following key issues will be addressed in each section. These issues are identified as referring to content (C) or organisational matters (O).

Developing the LSP test

- Establishing a clear need for the test together with the main stakeholder(s) and making it clear that the test will address mainly the language needs within a specific domain. (C)
- Data collection: Depending on the domain, it is probably necessary to observe, interview, collect evidence, and so forth. (C) During data collection sufficient reference materials and resources should be obtained. (O)
- Providing clear specifications, which include functional needs analyses. (C)
- Identifying the core tasks, lexis, functions and grammar for the test. (C)
- Relating (or not) the test to the CEFR or another external framework/standard, depending on the context. (C)
- Finding and involving suitable subject and context experts to work in partnership with throughout the whole test developing process. (O)
- Establishing an effective dialogue with stakeholders at every stage. (O)

Assembling LSP tests

- Recruiting and training suitable (expert) item writers. (O)
- Finding subject and context experts and assessment experts to review the tests. (O)
- Pretesting and trialling: Finding suitable candidates in sufficient numbers. (O)
- Content coverage: Ensuring suitable and sufficient coverage of the domain. (C)

Delivering tests

- Assuring live conditions for administering the tests or part of the tests suited to the type of LSP test (labs, classroom environment, etc.). (O)
- Assuring special types of input (specific equipment, special instruments or substances). (O)
- Using special facilities or tools (e.g., software). (O)

Marking, grading and reporting of results

- Recruitment and training of (expert) raters. (O) (C)
- Rating by a subject or context expert and an assessment expert (two persons), or an assessment and subject and context expert (one person). (O)

Monitoring and review

- Taking account of feedback from stakeholders in the expert domain, and feeding this back into development. (O) (C)
- Review should mirror development in the field. (C)

2 Developing the LSP test

2.1 The process of developing the LSP test

Most procedures of test development are identical for all language tests. Therefore, the test development team will consist mainly of experts of general language ability or proficiency tests (further called assessment experts). However, close collaboration with subject and context experts is crucial throughout the whole development process of LSP tests and should be taken into consideration from the start of test development.

2.2 The decision to provide a test

Authorities are often crucial stakeholders in LSP tests. In some cases, language requirements for the licence to practise in a (professional) field are defined by law. In other cases, there are professional bodies which set out rules or at least provide strong guidance with respect to the required language skills. In these cases, recognition of the test certificate depends on official approval. Therefore, the decision to provide an LSP test must be taken in close consultation with all stakeholders who have a voice in the recognition of the certificates.

2.3 Planning

2.3.1 Assessment experts

Assessment experts have the same responsibilities as in any other test development process, from needs analysis to the final delivery of results.

The role of the assessment expert in an LSP development project is to ensure that the highest technical and ethical standards are followed at all times. Another critical aspect of their work is to understand that the interaction with subject and context experts is critical. It is possible that the test developer may have a high degree of content expertise. However, this is unlikely to be the case on many occasions, so the test developer must be open to working with subject and context experts.

Assessment experts should keep in mind that professional subject matter is not their field of expertise. Even if the assessment experts feel that they have gained a good understanding of the subject matter in a particular domain, they should consult with individuals who have specific expertise in various aspects of the target area. It is also the case that this consultation is likely to be required throughout the development cycle in order to ensure that the tasks developed elicit appropriate language and also to ensure that they do not confound language and content.

2.3.2 Subject and context experts

The subject expert is a person with specialised expertise in a particular target area, e.g., an air traffic controller, a doctor. This person is critical in helping to define the test construct, i.e., the language and the production parameters of the context of language use, for example planning and speaking time. They can also help to identify sources of construct irrelevant variance, for example when a task is likely to result in language output that does not accurately reflect the construct.

The context expert might be a teacher of the language of a specific field (e.g., a teacher of English for Academic Purposes) or a professional who will interpret the expectations of the client or regulator (e.g., policy implementation specialist) in order to ensure that the resulting test is likely to reflect the needs of the client or receiving organisation(s).

2.3.3 Roles of the subject and context experts

When defining the test requirements for an LSP test the involvement of subject and context experts is crucial. Only they can shed light on the relationship between language and subject-related skills and the knowledge a candidate should master to perform adequately in a certain occupational field or profession. In most cases, these experts do not have a language testing or teaching background, which might make it more difficult to find common ground in all phases of test development. It is critical to allow sufficient time for consultations with these experts.

Subject and context experts will not always be able to describe language needs in specific linguistic terms. Therefore, for an LSP needs analysis, interviews with stakeholder groups should be conducted and, if possible, professional, and possibly other, stakeholders should be observed in their day-to-day work routines and contexts. These stakeholders can be divided into several groups (with nurses taken as an example here):

- colleagues working at the same professional level as the test takers after passing the examination, e.g., nurses in a hospital;
- colleagues in a different (higher or lower) position, e.g., doctors, cleaners;
- staff at the managerial level, e.g., the head of the nursing staff, administrative staff;
- academic/teaching experts, e.g., in the training of nurses;
- patients (and patients' close relationships).

When choosing subject and context experts, it is crucial to get the right mixture, with the first group (future colleagues) often considered as the most important one. However, the critical thing to remember is to look to all relevant stakeholder groups when undertaking the needs analysis.

It is advisable to question performance standards set by non-linguists. Almost everybody has an opinion about his/her own language and many subject and context experts believe that they can intuitively define what the right level of language skills is. However, these intuitive judgments may be skewed by non-linguistic factors or by a dated view on language use. They may also result in quite different opinions on the performance parameters and conditions, and the level of language required to undertake a specific role (see for example Berry, O'Sullivan, & Rugia, 2012). These different opinions need to be addressed by the assessment experts.

The quality and usability of the materials provided by subject and context experts, e.g., authentic texts, and the items they might write, always need to be checked by assessments experts.

2.3.4 A special group of context experts: LSP teachers

A very relevant group of context experts are LSP teachers, who bring in a complementary perspective. They know the domain, their learners, the questions and needs which learners bring to the classroom. Experienced LSP teachers can act as a moderator or interpreter bridging the views of professionals, their employers and the learners, and communicating this shared understanding to the test developers and assessment experts.

Linguistic domain

LSP teachers will be able to provide valuable linguistic insights in the TLU. However, they might show shortcomings in understanding the professional field at hand. Teachers, therefore, should not be the only expert group consulted.

Cultural domain

LSP teachers often have a good grasp of cultural differences which might cause difficulties when test takers begin to work in the field (see Japanese language test for tour guides in Douglas, 2000, pp. 176–178). This is especially the case if learners already have an internship during their language learning phase and encounter difficulties. It is, however, not easy to actually address intercultural matters in the framework of a formalised test. Test developers should consider which aspects are testable, e.g., the choice of register and forms of politeness. For example, there are clear differences between Japanese and American participants in business meetings in their interactional interpretations of silence and other nonverbalisations (Fujio, 2004). The question is not whether these differences need to be addressed in a test (clearly they should not be addressed in a language test), but how the language required to deal with such situations might be assessed.

Furthermore, it might be a good idea to focus not only on highly specialised language as found, for example, in a medical report. Adding a more general task might help to get representative samples of the test takers' competence besides testing their specialised language skills. For example, doctors learning a language might be quite able to write a medical report as there are many routines or formulaic sequences that can be learned and reproduced with relative ease. It is quite a different challenge, however, to write an informal note to a colleague. Experience shows that this kind of 'simple' task often presents test takers with difficulties as they are not yet able to use the target language well enough to adapt it to different genres or modes of communication that are part of a new workplace culture. LSP teachers are often familiar with these culturally determined conventions and expectations of both sides – the target workplace culture and the workplace culture of the learner. They are confronted with them in their teaching and can convey this information to the assessment experts.

Experience in the field

In some cases, LSP teachers have some work experience of their own in the professional field. This small group of experts is most valuable. The LSP teachers can provide valuable input on linguistic functioning. They are also well suited to write items themselves, provided that they are trained. Items should always be checked for authenticity by professional practitioners.

If possible, both subject and context experts (e.g., LSP teachers) should be consulted in the initial process of a needs analysis, in the development of test specifications and test items. The subject and context experts have a consulting role, providing their perspective on the relevance and authenticity of test tasks.

The three main groups of experts in summary:

Experts	Examples of roles
Assessment experts	Overall test development experts, project managers, trainers of item writers, examiners, raters
Subject experts	Informants about the domain requirements, consultants on construct definition and in test item/ task development, co-trainers of item writers, providers of authentic material, item writers, reviewers of test items
Context experts	Informants about specific language needs related to the domain, consultants about intercultural differences, item writers, reviewers of test items

It is important to remember that other experts may need to be consulted. For example where specific national or international regulations are in place (e.g., air traffic control), experts in the legal and professional application of these regulations will contribute to the test developers' awareness and understanding of the context in which the test will be used.

We can see, therefore, that an LSP test is very much a collaborative process, with the assessment expert on the one hand, and the subject and context experts on the other hand, working together throughout the process.

EXAMPLE: Informants for a test for nurses

- **Subject experts:** Nurses working at a hospital (if that is the target workplace) can provide valuable insight into the linguistic challenges of the job. Interviewing and/or observing them can lead to insights in, for example, the kinds of texts nurses have to write, the typical format of specific texts, the time frame in which a text of a certain length has to be produced, and the needs of the audience. Further analysis of typical content and linguistic features has to be conducted by an assessment expert. Consultation with doctors can result in understanding the specific collaborative tasks between these two professional categories. Test tasks could be elaborated to include, for example, the need of the candidate to understand the indications and instructions given by a doctor and to act accordingly.
- **Context experts (teachers):** Teaching nurses in LSP classes, teachers will have developed an understanding about typical text features. They often work with authentic samples from their learners' future workplace (maybe they are even teaching at the hospital itself) and they know how to teach writing and speaking skills. They will also be aware of interferences from other languages, in addition to cultural biases.
- **Health regulators/Policy makers:** Individuals who are expert in national/regional policy and policy implementation can advise on how policy makers/clients will perceive specific aspects of the proposed test, in particular tasks and reporting systems. They can also advise on validation approaches – it is unlikely that an assessment argument designed to appease test theorists in the USA (e.g., a Kane-based assessment use argument) will satisfy the expectations of a German or Irish nursing regulator who will most likely have specific concerns that should be addressed in a validation argument.

2.3.5 Needs analysis

Before putting together expert informants for a needs analysis, the assessment expert should first consult the literature on the teaching, learning and assessment of the language and content of the target area. This might include academic/scientific literature, technical reports and more practical documents such as subject-specific language can-do statements (where they exist). Other test developers (and experts) might also be consulted in this phase. Having gathered a group of experts, the process of needs analysis begins. Crucial questions for LSP test development are:

- Are there official regulations or laws which define the language test or test delivery conditions in any way?
- What is the social dimension of the test?
- What is the test's likely impact? Tests that give admission to professions or academic admission tests are high-stakes, often more so than general language tests.
- If the certificate is used for access to a profession or to academic studies, at what stage of the learner's journey towards his/her goal is the language test administered?
- Which types of tasks best reflect the needs of the workplace? How authentic do the tasks need to be? How similar to the actual communicative tasks in the workplace?

- How do the language demands interact with the professional subject knowledge and skills in the field?
- What general language skills are needed in the TLU domain that the LSP test targets (i.e., a CEFR level in particular skills, in all skills, etc.)? Is there a threshold level? How has this been set?
- What specific language skills are needed for functioning in the workplace? Which domains, topics, themes and social contexts are relevant?
- What are the lexical, functional and grammatical requirements of the language used in the workplace?
- What role does jargon, technical language and professional language play?
- To what extent does cultural knowledge need to be incorporated into the test? How will this be done?

A needs analysis may include:

- consultations with subject and context experts;
- identification of any other critical stakeholder groups and inclusion of these in the process;
- surveys for experts as specified above, as well as learners/potential test takers;
- interviews with practitioners and/or different categories of stakeholders;
- curricula or teaching programmes;
- authentic written and oral materials: learning materials, tasks, texts, textbooks, recordings of professional activities;
- systematic observations at the workplace.

EXAMPLE: Questionnaires for learners/potential test takers

In a test commissioned to assess hotel industry language use, the information that emerged from the needs analysis questionnaire showed that the majority of the test takers would function as front-of-house staff, and that they would need to interact with customers. In this case, the test will need to focus more on the oral skills needed in staff–customer interactions.

2.4 Design

2.4.1 Initial considerations

All test design starts with a clear understanding of the test-taking population. When an LSP test is first defined, experts often find it useful to imagine the test taker in his/her (future) work environment, e.g., on the runway of an airport. One should, however, keep in mind that an LSP test, even if it is highly specialised, is still a language test. It is, therefore, central for the test construct to elicit relevant and authentic communicative speech acts. This aim is not necessarily accomplished by testing at the workplace or by using instruments which are needed for the job. For the information gathering process there are several possible sources which will help to understand the test content and format as well as the TLU in the domain: literature, textbooks, (language) learning resources, job descriptions, but above all experts in the field.

Therefore, the test construct (and the tasks derived from it) need to reflect closely the TLU domain, which in the case of an LSP test can be very specific.

When scoring and rating criteria are discussed, the relation between content specific (e.g., task fulfilment) and more general linguistic (e.g., grammatical correctness) criteria needs to be clarified – keeping in mind that the ‘perfect’ candidate probably does not exist. What needs to be determined are the language skills of the ‘minimally acceptable person’ (see the *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): A Manual*, Language Policy Division, 2009, p. 62). This is someone who can cope with the job at hand in a sufficient manner, although a higher level of competence might be preferable.

EXAMPLE: Police

When tests are developed for a particular organisation, often that organisation will describe a perfect person rather than a minimally acceptable one. For example, the perfect police officer will be completely fluent and able to communicate in all kinds of language varieties and registers.

Expecting advanced, C2 skills would exclude most applicants, and might disregard the question of the actual language performance of L1 speakers on the job or what language level professionals have to master in order to perform their duties in an adequate way. Therefore, it is necessary to determine the linguistic threshold below which a person’s proficiency

performance becomes inadequate.

2.4.2 How to balance requirements with practical considerations

Authenticity

The necessary degree of authenticity of the test tasks must be determined. Is it necessary or desirable that candidates perform exactly the same task which they have to do on the job (e.g., a guide giving a tour or a pilot flying a plane)? Or, is it sufficient to identify the main language functions and to elicit them in an examination environment which could be more distant from the actual work context? Note that this discussion cannot be made by the test developer alone. All authenticity issues must be dealt with in consultation with subject and context experts.

Impact

In the process of reconciling requirements and constraints, it is crucial to discuss what information a test can actually provide. Stakeholders outside the language community need to accept that a language test alone cannot predict success on the job, at university, etc. as there are too many other factors apart from language that play an important role here.

Practicality

Depending on the specificity of the test, a more extensive cost–benefit analysis might be needed than with general language testing. A general first approval by the relevant authorities does not mean that the new LSP test will be a success; financial sustainability needs to be taken into consideration. First of all, the target group of test takers is always smaller than for general language ability tests. In addition, it is time-consuming and therefore expensive to conduct a thorough needs analysis in a specialised field and to involve subject and context experts (which you do not need for general purpose tests).

If a relatively low number of test takers is expected, this has an impact on test development. It is necessary to find institutions where pretesting and trialling can take place. If not enough quantitative data can be collected, it is crucial to plan on sufficient qualitative feedback in order to ensure quality.

2.4.3 Test specifications

The process of designing LSP test specifications is similar to that involved in general purpose language tests.

We recommend that the test developer always uses a model of test development and validation when creating the specifications as this tends to ensure that all relevant issues are systematically dealt with during the process. One useful approach is provided by the socio-cognitive model (Weir, 2005). See O’Sullivan and Dunlea (2015) for a detailed example of a test that has been developed using the model.

2.5 Try-out

Additional aspects to be considered in the try-out phase of the LSP test development are:

- Are there enough representative test candidates? For some LSP tests a qualitative, small-scale try-out or piloting might be the most relevant option, especially in the case of very specific tests, such as admission tests for a specific profession. In such cases, representative native speakers can also be included in the piloting of the LSP test (e.g., those in training for the job).
- In case of very few candidates, take care not to pilot the test on future test takers.

Generally, at every step of test development, it is important to check with the stakeholders whether the chosen task types represent common practice in the aimed profession (i.e., qualitative feedback).

2.6 Informing stakeholders

Some stakeholders, for example employers, might be critical of what they see as assessment experts judging professional, non-linguistic competencies. On the other hand, these stakeholders may feel that they are able to judge the language competencies necessary to function at the workplace. In order to avoid misinterpretations, a dialogue between all parties involved has to be established. It might be productive to organise this dialogue around the criteria for rating test takers’ performances, task fulfilment or other aspects relating to language assessment literacy (see also section 3.2.1).

For all stakeholders it is essential to receive information on how the LSP tasks were developed, in what way subject and content

experts were involved and how the rating process and other validity issues were addressed. In other words, stakeholders should be considered when developing validation reports/arguments. See O'Sullivan (2016) for a detailed discussion on the role of stakeholder groups in the test development and validation process in which these key groups are seen as central to the whole process, from the initial construct definition to the construction and delivery of any later validation reports.

2.7 Key questions

- Who are the stakeholders?
- Who decided that a test should be developed? Are there any official regulations?
- How will the test results be used in the domain/sector?
- How can the quality of the test be assured? What are the minimum requirements referring to quality?
- How many test takers will be tested? If a relatively low number of candidates is expected, which provisions are being taken for quality assurance (in the try-out phase as well as later on)?
- What insight into the test use and the requirements of various stakeholders can subject and context experts provide? To what extent can they define the required language skills?
- What information can observations of and interviews with stakeholders provide?
- What are the core lexis, functions and grammar in the work context or profession?
- What types of tasks are necessary to elicit the required language skills?
- Do the chosen task types represent common practice in the profession?
- How can practical constraints (time, finance, etc.) and stakeholders' requirements be balanced?
- How can stakeholders in the field be informed appropriately?

3 Assembling LSP tests

3.1 The process of assembling tests

The process of assembling an LSP test does not fundamentally differ from the steps that have to be taken for general language tests.

3.2 Preliminary steps

3.2.1 Item writer recruitment and training

In Chapter 2, different kinds of experts and their role in developing an LSP test were discussed. In the assembling stage, the subject and context experts (especially LSP teachers) might be recruited and trained as item writers. If these experts write items, assessment experts have to check whether the items work as intended with respect to language. On the other hand, when the items are written by experienced assessment experts, the authenticity of the tasks should be checked by a subject and/or context expert.

A good way to divide roles between the two main groups of experts might be to let the subject and context experts find authentic texts and contexts as a starting point, whereas the assessment experts might be more suited to write the actual items.

Whatever the exact role for each group of item writers, they all have to be trained adequately. The cooperation works best if:

- the subject and context experts understand that their intuitive understanding of language requirements might not enable them to write items as needed for a language test;
- the assessment experts accept that – even though they might have gathered considerable content knowledge in the field – they might not be able to judge the finer points in the professional field.

Both groups of experts need to understand each other's perspective and to follow the test requirements as laid out in the test specifications. One should be aware that the interaction between all the experts might take up a lot of time and might be challenging for the participants.

Overview of the experts' roles and characteristics in the item writing process:

Experts	Roles	Characteristics
Assessment experts	Monitoring and supervising of the item writing process (recruitment, training, reviewing, etc.), item writing, item/task reviewing	Have previous experience in designing tests for general language ability or language proficiency, acquire up-to-date knowledge in the LSP field
Subject experts	Training of assessment experts in content domain requirements, providing texts/sources for the test items, supporting item writers in understanding certain materials (texts, areas of vocabulary, etc.) in TLU situations, item writing, writing of samples of anticipated response for items, item/task reviewing	Have current experience in the domain, can detach themselves from their specific workplace and take a broader perspective on the field
Context experts: LSP teachers	Training of subject experts in content domain (linguistic) requirements, item writing, item/task reviewing	Have expertise in both language and the subject content

For understanding and being able to use content-specific input, an item writer needs to familiarise themselves with the language of the specific field, e.g., terminology, rhetorical structures, etc. (Selinker (1979) in Douglas, 2000, pp. 97–98).

Some methods for assessment experts to acquire subject expertise are:

- train with the help of subject experts;
- have preparatory stages, in which the items they produce are assessed by experienced LSP item writers or subject expert item writers and feedback is given;
- consult vocabulary lists, literature and authentic (written) material in the field;
- familiarise themselves with the specific domains (e.g., by grounded ethnography, context-based research – Douglas, 2000, pp. 93–96) and discuss the information obtained with subject experts;
- audit or take an LSP test in the domain, previously produced by experienced LSP item writers, in order to familiarise themselves with the specific language, the specificity of the tasks and the perspective of the test taker in the actual target language contexts.

3.2.2 Managing materials

The professional domains the items are created for could be highly specific, quite varied and not always easily and clearly distinguished from one another for a non-specialist, so item management might be a considerable challenge in terms of content. Therefore, subject experts need to be consulted in the process of categorising these items according to the professional domain addressed in the test.

EXAMPLE: A test for nurses in hospital and in nursing homes for the elderly

If, for example, a test for nurses is designed to cover two workplaces, i.e., hospitals as well as nursing homes for the elderly, it is not always easy to accommodate both areas equally well. While the practitioners of geriatric or hospital care may feel their respective fields to be quite well defined and easily distinguishable from one another, assessment experts might well miss the finer differences. Hence, they might not be able to categorise items correctly and consequently might get the mix of tasks wrong.

3.3 Producing materials

The cycle of the production of items closely follows the one for general language tests, with the potential involvement of specialists at all stages.

3.3.1 Assessing requirements

It is always a relevant question how many items need to be produced in order to have a functioning item bank which covers the need for a certain period of time. If an LSP test is either relatively new or if there are few test takers, item writers might not have the same routine in producing fitting items as in the case of general proficiency tests. Commissioning more items than the minimum necessary is highly advisable. A certain surplus (at least 30%) is needed as more items than usual might have to be rejected after pretesting. Two ways to improve on the efficiency of item writing are:

- to have very clear written specifications (and item writer guidelines);
- to build expertise amongst the item writers by asking individuals to focus on particular types of tasks or items until they gain sufficient experience.

3.3.2 Commissioning

The list of requirements (details of the materials required by the commissioner) should specify as clearly as possible what the item writers' and reviewers' tasks are. For LSP tests, the regular experienced item writers might not be qualified. Therefore, the test provider will have to find new LSP item writers and subject experts who might not have much experience in producing materials for assessment purposes. This does not only relate to the content of test items, but also to formal aspects (e.g., keeping to the specifications in terms of text length, maintaining an adequate register, selecting visual prompts).

Apart from the test specifications, helpful documents for item writers include:

- a profile of the target test taker;
- a list or lexicon defining the range and level of vocabulary and/or language structures to be used;
- relevant sources of material or texts of reference (e.g., domain-specific dictionaries or encyclopaedias) for the domain.

Where available (i.e., in an existing test), the following may also be of use:

- sample student performances;
- past papers, marking criteria, answer keys.

3.4 Quality control

3.4.1 Editing new materials

In the case of LSP tests, the process of checking the quality of the materials has two dimensions:

- the language aspect (CEFR level, language functions, grammar, vocabulary, etc.);
- the specific purpose background knowledge (whether the materials are accurate, whether they ensure suitable and sufficient coverage of the domain, etc.).

Items should be checked by two types of specialists: assessment experts on the one hand and subject and context experts on the other hand. However, the participation of subject and context experts in editing meetings is likely to be limited to the early stages of the development and delivery of the test. It is highly unlikely that such experts will engage with the test beyond this stage due to cost and personnel resourcing issues.

During editing meetings, new item writers should have the opportunity to participate in checking the quality of draft materials and to learn from experienced item writers and subject and context experts.

In case there are constructed response or open-ended questions, the key must provide more information than in the case of selected response items: the rater will not only need very clear and precise definitions and examples of (task-specific) characteristics that are to be considered in assigning a score, but also guidance on how to deal with factual, content-related errors.

3.4.2 Piloting, pretesting and trialling

Piloting, pretesting and trialling or a combination of these methods must be used.

However, pretesting and trialling, which always require a certain number of participants, can be difficult for LSP tests. There is a smaller number of LSP test takers in general, so finding suitable candidates for pretesting and trialling in sufficient numbers could present a problem.

Therefore, piloting the materials might become very important. In a pilot, qualitative feedback on both content and language should be collected from item writers specialising in the same domain, students involved in the area of interest for the test, and subject and context experts. Feedback might be given by both native and non-native speakers of the target language.

3.4.3 Review of items

The procedures in the review phase following piloting/pretesting do not differ from those in general language testing. However, given the specificity of the domain(s) for which the test is created, essential factors involved in the process of piloting/pretesting (e.g., number of test takers and/or their representativeness for the target population) might not produce the most relevant results in statistical analysis. In consequence, qualitative data need to be collected from test takers, test providers, experienced item writers and content and context experts.

3.5 Constructing tests

Once enough materials are collected, tests can be constructed.

The test construction stage involves balancing a number of different aspects, such as test content and item difficulty. Assessment experts as well as subject and context experts should be involved in this process. For example, it might be difficult to establish item difficulty solely based on quantitative analysis. Qualitative input from both types of experts might be needed to determine item difficulty.

3.6 Key questions

- Who will write the materials: assessment and/or subject and context experts? What should the specific requirements for item writers be, particularly with respect to content?

- How will the assessment experts who write items be trained in the specific professional domain? Who will carry out the training?
- How will subject and context experts who write items be trained in the language requirement of the domain?
- How will topical content, authenticity, context, and the other variables be balanced in item writing?
- How comprehensively is the language of the domain defined or described and how is it reflected in the items?
- To what extent is the help of subject and context experts necessary? What is going to be the role of these experts in the test assembly phase?
- How will the items be piloted, pretested or trialled? Are there enough test takers?
- What type of analysis is more adequate for the performance data gathered through piloting, trialling and pretesting, especially if there are not enough candidates for these?
- Are subject and context experts going to be involved in editing meetings?

4 Delivering tests

4.1 Aims of delivering tests

In general, an LSP test is delivered as any other language test, either as a paper and pencil test or online. Testing in the working environment could be a special LSP feature, for example, if an air traffic controller were tested in the control tower at the airport.

4.2 The process of delivering tests

Most people interested in an LSP test will already be active in the labour market. They may be satisfied employees who are interested in a language qualification as a route to promotion, or persons who might intend to find a job or change jobs. Alternatively, they may be an employer who is looking for help in selecting appropriate staff. As with any test, administration should be organised to fit the agendas of the test users or takers, with tests taking place over a weekend or a holiday period if required. This may have an impact on the registration process and its deadlines.

4.2.1 Arranging venues

It is not uncommon for an LSP test to take place at the workplace. Some employers organise in-house language training and want to offer the test at the workplace as well. In these cases it is important to ensure the confidentiality of testing materials. Great care has to be taken that test takers can take the examination under standardised conditions in order to prevent fraud and ensure a fair examination.

If a trusted person is charged with the inspection of the venue, attention must be paid to possible conflicts of interest. Some groups of professionals are small, so they all might know each other (e.g., coroners in Belgium, maybe 40 people maximum), which might put the trusted person in a conflicted position. For objective quality assurance, the inspection of the venues (and of the whole test administration process) should be carried out by employees of the testing institute.

4.2.2 Registering test takers

In LSP contexts more information about test takers may be useful in order to monitor test use and to plan revisions as necessary. The test provider should consider the need to gather the following information about the test takers:

- learning experience (in a general language proficiency course or an LSP course);
- professional background (professional training, work experience, work experience in the target language);
- purpose for taking the test;
- relevant personal data, like L1.

If test takers with special needs register, it is advisable to consult subject and context experts in order to determine what kind of special arrangements are usually made in the workplace (in addition to standard test accommodations). It is advisable to arrange for special needs in accordance with the most common accommodations taken at the workplace.

4.2.3 Sending materials

The same procedures as for general language tests apply.

4.2.4 Administering the test

With respect to test takers' behaviour during the examination, test takers should know in advance which specific 'supporting materials', if any, they are allowed to use. For example, if specialised dictionaries (e.g., containing professional terminology) are allowed, which particular dictionaries are allowed, and should the test takers bring their own copy or do they have access to a number of them in the test venue?

4.2.5 Returning materials

The same procedures as for general language tests apply.

4.3 Key questions

- How could the test administration dates be adjusted to the test takers' work schedule, seeing as they might already be active in the labour market?
- Are subject-specific supporting materials allowed during examinations and are they available at the venue?
- What test accommodations are needed for test takers with special needs?

5 Marking, grading and reporting of results

As in all other phases of LSP assessment, the main difference between a general purpose and an LSP test at the marking and grading stage can be captured in the question of whether or not subject and context experts should be involved.

5.1 Marking

5.1.1 Clerical marking

As clerical marking is a highly standardised job, it is not necessary that markers understand the technical jargon of the field.

Guidance on evaluating responses

Special attention needs to be given to the key in LSP tests. As for any other test, it should contain all the alternatives. However, in the case of an LSP test, these should be agreed on by the development team and subject and context experts (not just a single person). This is to ensure that all viable responses are acknowledged. Again as in all tests, markers should be asked to record and report all alternative answers they find which they feel might be correct. With LSP tests, markers should be especially careful in the sense that alternatives may sound plausible for an assessment expert, but may not reflect authentic language use in the context of the specific domain. The alternative answers that come up during the process of marking, for example in a gap-filling task, should be evaluated by both assessment and subject and context experts in order to reach a decision on whether the alternative is correct or not. Therefore, one subject or context expert should be available as consultant during the process of marking/grading. This is particularly important at the early stages of marking, and practicality means that a line must be drawn under the list of acceptable responses as early as possible since constant updating means that all previously marked responses must be re-visited and re-marked. Where an online automated marking system is in use, the situation is less problematic, as the updated list of acceptable responses can quite easily be applied retrospectively at the end of the marking event.

One example of why it is necessary to engage with subject and context experts is the very nature of specific purpose language. Use of language for specific purposes is more precise and accurate than general language (Douglas, 2001). The more specialised a domain is, the fewer synonyms are usually possible. The typical marker/examiner may not be aware of the importance of particular words or expressions within a given LSP context. This is critical in situations where it is crucial to ensure absolutely unambiguous communication, for example for pilots or train drivers.

Managing the process of marking

Even if the number of LSP test takers is usually smaller, this does not mean that the process of marking is necessarily easier to organise. Either the exam is rarely administered, in which case the markers cannot acquire routine; or there are many exam dates with only very few candidates or even a single candidate (who needs the certificate to get a job as soon as possible).

In many LSP contexts there are time constraints generated by stakeholders or beneficiaries (companies, educational institutions, professional bodies, etc.) which might have consequences for the marking and the rating process.

5.1.2 Machine marking

Machine marking for LSP tests is similar to those of other language tests, though see the note above on identifying acceptable responses using subject and context experts.

5.1.3 Rating

Both assessment and subject and context experts need to be involved in rating. The key for constructed response or open-ended questions, for example, should be designed by subject and context experts and assessment experts together.

Rating productive performance in LSP tests might pose greater challenges than general language tests. In the case of highly specialised tests it might be useful to involve subject and context experts as well as assessment experts. The necessary degree of cooperation might depend on the degree of specialisation of the test and on the degree of jargon typical of the subject area. However, it is important to remember that this is primarily a language test, so the likelihood of having subject and context experts available to rate performances might be low. Instead, we should look beyond the traditional involvement of such experts in rating scale development to involve them in rater training (design and delivery), and in monitoring rater behaviour (during or post rating).

There are a number of different forms of cooperation that can be implemented, for example:

	Advantages	Disadvantages
The subject/context expert as the first and decisive rater, the assessment expert as the second rater	Clear role definition	Subject/context expert may focus overly on content rather than language (i.e., introduce bias)
The assessment expert as the first and decisive rater, the subject/context expert as the second rater	Clear role definition	Potential for strong second rater to overrule assessment expert
Both experts rating together and reaching the result by consensus	Usually easy to come to a decision	Criteria may change from performance to performance Rater personality may allow one to dominate Can result in difficult-to-resolve disagreements
Both experts rating independently, using different criteria (holistic and analytic)	Theoretically the most solid approach – each looking at slightly different aspects of the performance	Difficult to balance the two views to get a fair estimation of ability
The subject/context expert checking the assessment expert's rating	Offers a systematic double rating	The two raters may be looking at different aspects of the performance. This can lead to confusion and inconsistency
The subject/context expert acting as a trainer and consultant	Brings consistency of interpretation of the rating scale and of performance level Helps ensure consistency of interpretation across a rating event (where used for monitoring)	Potential to influence the assessment expert focusing too much on content at the expense of language aspects

In order to determine the most appropriate approach to take, the test developer should consult with test users and with the test sponsors, e.g., where a professional organisation commissions a test. Since engaging with subject and context experts can be difficult in terms of availability and cost, the decision is generally made on a practicality basis. This was certainly the case with the OET in Australia, where such issues were identified early in the development process (McNamara, 1990). The most important thing to remember, as with any test, is what claims will be made based on test performance. If we are trying to establish whether an individual has sufficient and appropriate linguistic resources to engage with a specific work or study domain as a professional or as a learner, then we are making language claims. So, the test we develop must focus primarily on language. There will always be a question mark over the ability of assessment experts to make decisions related to a specific work or study domain due to the vagaries of language use across different domains. The onus is therefore on the test developer to base all decisions on what the test focuses on and to offer a systematic and defensible rationalisation for all decisions made during the development phase. Concerns can be alleviated by then communicating this rationalisation openly and efficiently to test users and sponsors. This is what validation is all about.

5.1.4 Rating criteria

When developing a rating scale for use in an LSP test, it is critical that the criteria which make up the scale are clearly and explicitly stated, as is the case with any language test. It is also critical that they are appropriate to the language of the domain being tested. If the test is to focus only on language, then the criteria and contents should be agreed on and defined by both subject/context experts and assessment experts working together. Where the test is designed to integrate the language and the content more robustly, as would be the case in a highly specific test, then clear lines should be drawn between criteria that focus on language and those that focus on content. See Abdul Raof (2011) for how this can be achieved in a highly specific test.

The most important lesson to learn here is that for a scale to work appropriately it should be designed and developed by both subject/context experts and assessment experts. It should then be exemplified by performances chosen by these experts and any training programme should be devised and, where possible, delivered by both.

EXAMPLE 1: Criteria to be judged either by subject/context experts or by assessment experts in an LSP test for nurses

It is necessary for a nurse, for example, to speak clearly as he/she must be understood by elderly patients whose hearing is not good. This need can be identified by professional experts and clearly described in the rating scale and illustrated in training. However, judgment about pronunciation is and should be the domain of the assessment expert, though clearly the subject/context expert will provide input on the level or levels of acceptability (see row two in the table in section 5.1.3).

EXAMPLE 2: TEACH test (Taped Evaluation of Assistants' Classroom Handling)

Among the criteria for rating communication skills are: teacher presence (appearance, attitude), indication of his/her awareness of the appropriate teacher–student relationship in a US university classroom setting, showing interest in the subject and in the students as learners (Douglas, 2000, pp. 163, 251) (see row four in the table in section 5.1.3).

As with any language test, assessment experts need to be trained before rating. This is even more critical with LSP tests than with other more general language tests as they will need to be familiar with the scope of language use specific to the LSP domain and to understand the use of language in oral and written discourse in terms of precision of use, variety, breadth, etc. Also, in certain LSP domains, such as the professional one, written communication often aims at clarity as well as at fast absorption of the salient facts. Therefore, the linguistic variation usually expected at high CEFR levels might not occur in authentic texts. This has to be taken into account if raters of general language are trained to judge LSP performances. However, it is only possible to identify appropriate linguistic variation by engaging with subject and context experts.

5.2 Grading

In grading, there is no significant difference between general language and LSP tests. However, it is advisable to include, where possible, subject and context experts in the process as advisors/consultants. This is certainly the case in early administrations of a test when it is possible that the systems are not fully functional and some interpretation of appropriateness or adequacy of performance is still required.

5.3 Reporting of results

The general rule that results should be clear and meaningful for stakeholders who do not have much knowledge about language exams, CEFR scales and the like applies to all language certificates. For LSP tests, it is advisable to make very clear that the certificate relates to language competence, not to professional competence. There should be no ambiguity in this respect in order to avoid unintended uses of the certificate.

5.4 Key questions

- Will both assessment experts and subject/context experts be involved in the training of item writers?
- Will both assessment experts and subject/context experts be involved in marking and rating? What exactly will their role be?
- What is the relationship between the language and the subject criteria in achieving the final results?

6 Monitoring and review

For LSP tests monitoring and review are constant processes of quality management as for any other language test.

Even though the overall suitability of the test format for the purpose will have been ascertained in the course of test development, the test providers need to keep in touch with stakeholders from the field in order to remain informed about any major technological or procedural changes in the field. If these changes or developments have an impact on language use, a modification of the test might be necessary. It is advisable to work with an informant group from the domain on a regular basis in order to be certain that developments in the domain are monitored. An annual meeting or report is usually sufficient for this purpose, though the developer should always consider building in a regular review period for any language test. Test performance data can be gathered routinely over all administrations and a formal review project undertaken, for instance, every five years (O'Sullivan & Dunlea, 2015). This helps to guarantee the viability and long-term validity of the test.

It is useful to carry out predictive validity studies on the test takers' long-term linguistic success in the field. In this context it is important to differentiate between language skills and overall subject-related competence. If a doctor prescribes the wrong treatment or if a pilot crashes, the reasons will usually lie in their professional skills. A language certificate can neither predict overall competence on the job or general suitability for a career, nor factors such as interpersonal skills which contribute to a successful integration into a work environment. However, it is useful to find out how test takers cope linguistically and whether, or to what extent, the test score predicts their linguistic integration at the workplace. This kind of study is a good starting point for a periodical review of the test.

Another aspect to monitor is test use. While the test might be designed for a certain occupational field, e.g., nurses, stakeholders might want to use it for related fields as well, e.g., care givers for the elderly, or for a completely different purpose. While this unintended use of test results often is beyond the control of the test developers, test use should be monitored and stakeholders advised. This is a controversial area, with test developers arguing that they cannot be held accountable for how their tests are used in all instances and theorists arguing that they have an ethical responsibility to their test takers to ensure that their rights are not undermined (e.g., by unsubstantiated claims being made about their ability to perform, or not, linguistically in a domain not related to the original focus of the test). It is clear that test developers should make their position on test use clear to test users and lobby against particular uses where these are deemed inappropriate.

6.1 Routine monitoring

Due to comparatively low numbers of test takers, it often takes a long time to get results from monitoring, be it qualitative or quantitative. Changing parts of the test in reaction to feedback from small groups or even individual stakeholders is not advisable. The test provider should refer critics back to the initial needs analysis and development of test specifications. Only if there is substantial demand for modifications, i.e., well supported by facts and brought forward by several independent stakeholders, should a change be taken into consideration outside of the routine review policy.

The test provider may want to remember that each stakeholder has their specific (work) environment in mind when commenting on test specifications and results. Depending on the specificity of the LSP test, stakeholders may not be aware that a language test has to generalise with respect to language competence and that it does not aim at reflecting a specific workplace. Test providers may want to communicate to the stakeholders that the communicative needs of the workplace might be captured in a test without recreating the exact workplace.

A review should be made if new regulations are introduced by law or by relevant authorities. If, for example, the end user is required to prove a higher level of language proficiency than before, the original test has lost its value. Another example can be a changing syllabus for a qualification in the field. In such a case, a modification of the test might only be necessary if this kind of change actually affects communicative needs. A new syllabus may introduce new specialised vocabulary, which the test developer may need to incorporate into the test specifications.

6.2 Periodic test review

Both assessment experts and subject/context experts should be involved in the review process.

6.3 What to look at in monitoring and review

The same procedures as for general language tests apply.

6.4 Key questions

- What data needs to be collected from which stakeholders in order to monitor and review the test effectively?
- In which proportion should assessment experts and subject/context experts be involved in the process of test review?
- What changes in the professional field call for a review of the test?

References and further reading

- Abdul Raof, A. H. (2011). An alternative approach to rating scale development. In O'Sullivan, B. (Ed.), *Language Testing: Theories and Practices* (pp. 151–163). Oxford: Palgrave.
- Alderson, C. J., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Berry, V., O'Sullivan, B., & Rugea, S. (2012). Identifying the appropriate IELTS score levels for IMG applicants to the GMC register. London: University of Roehampton/The General Medical Council.
- Brown, A. (1993). LSP testing: The role of linguistic and real-world criteria. *Melbourne Papers in Language Testing*, 2(2), 35–54.
- Brunfaut, T. (2014). Language for Specific Purposes: Current and future issues. *Language Assessment Quarterly*, 11(2), 216–225.
- Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Davies, A. (2001). The logic of testing Languages for Specific Purposes. *Language Testing*, 18(2), 133–147.
- Douglas, D. (2000). *Assessing Language for Specific Purposes*. Cambridge: Cambridge University Press.
- Douglas, D. (2001). Language for Specific Purposes assessment criteria: Where do they come from? *Language Testing*, 18(2), 171–185.
- Fujio, M. (2004). Silence during intercultural communication: A case study. *Corporate Communications*, 9, 31–39.
- Hackett, E. (2002). Revising the BULATS Standard Test. *Research Notes*, 8, 7–9.
- Ingham, K. (2008). The Cambridge ESOL Approach to Item Writer Training: the Case of ICFE Listening. *Research Notes*, 32, 5–9.
- Ingham, K., & Thighe, D. (2006). Issues with developing a test in LSP: the International Certificate in Financial English. *Research Notes*, 25, 5–9.
- Language Policy Division (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment CEFR. A Manual*. Strasbourg: Council of Europe.
- McNamara, T. F. (1990). *Assessing the second language proficiency of health professionals*. Unpublished Ph.D. thesis, University of Melbourne, Australia.
- O'Sullivan, B. (2002). Some theoretical perspectives on testing language for business. *Research Notes*, 8, 2–4.
- O'Sullivan, B. (2006). *Issues in Testing Business English: the Revision of the Business English Certificates*. Studies in Language Testing Volume 17. Cambridge: UCLES/Cambridge University Press.
- O'Sullivan, B. (2012). Assessment Issues in Languages for Specific Purposes. *The Modern Language Journal*, 96, 71–88.
- O'Sullivan, B. (2016). Validity: What is it and who is it for?. In Leung, Y. (Ed.), *Epoch Making in Teaching and Learning: A Special Monograph for Celebration of ETA-ROC's 24th Anniversary. Selected Papers from 2016 PAC & The 25th International Symposium on English Teaching* (pp. 157–175). Taipei: Crane Publishing Company Ltd.
- O'Sullivan, B., & Dunlea, J. (2015). *Aptis General Technical Manual: Version 1.0. TR/2015/005*. London: British Council.
- Paltridge, B., & Starfield, S. (2013). *The Handbook of English for Specific Purposes*. Malden, MA: Wiley-Blackwell.
- Selinker, L. (1979). Interlanguage. *International Review of Applied Linguistics*, 10, 209–231.
- Smirnow, S., Maes, C., Luyten, L., & Verheyen, S. (2017). Assessment in a globalized economy: A task-based approach to assess the proficiency of Dutch in specific occupational domains. In ALTE (Eds.), *Learning and Assessment: Making the Connections – Proceedings of the ALTE 6th International Conference, 3–5 May 2017*, (pp. 26–31). Retrieved from <http://events.cambridgeenglish.org/alte2017-test/perch/resources/alte-2017-proceedings-final.pdf>.
- Thighe, D. (2007). Cambridge ESOL and tests of English for Specific Purposes. *Research Notes*, 27, 2–4.
- Weir C. (2005). *Language Testing and Validation: An Evidence-based Approach*. Basingstoke: Palgrave MacMillan.
- Wu, W. M., & Stansfield, C. W. (2001). Towards authenticity of task in test development. *Language Testing*, 18(2), 187–206.

Acknowledgements

Editing team: Koen Van Gorp and Dina Vîlcu
with the support of Ina Ferbežar and Sibylle Plassman

The editing team wishes to acknowledge the contribution made by all members of the ALTE Language for Specific Purposes Special Interest Group, and especially the following members for producing and reviewing draft texts on several occasions:

Kate Biggins
Dominique Casanova
Rebecca Chick
Paul Crump
Etienne Devaux
Linda Fromme
Marta García
Jelske van der Hoek
Sien Joos
Wassilios Klein
Hanne Lauvik
Petra Likar Stanovnik
Lucia Luyten
Gemma Macho Aguillo
Helga Christine Maes
Andrea McKay
Florian Nimmrichter
Francesca Parizzi
Graham Seed
Annika Spolin
Steven Verheyen
Naomi Withmore

Reviewer: Barry O'Sullivan

Production Team: Alison French
Mariangela Marulli
Jane Rust
John Savage
Servis Filmsetting

