ASSESSMENT: A CHALLENGE FOR ESP PRACTITIONERS

La evaluación: un reto para el inglés para fines específicos

Sonsoles Sánchez-Reyes Peñamaría* y Gabriela Torregrosa Benavent**
Universidad de Salamanca
Correo-e: *sreyesp@usal.es; **Gabriela.torregrosa@usal.es

ABSTRACT: LSP tests are having a growing demand nowadays, especially business English. LSP tests comprise academic and occupational or professional purposes. This article shows an overview of the most important aspects LSP practitioners have to take into account when preparing LSP tests, as they are different from general language ones. LSP tests should be valid to measure the learner's professional needs in terms of authenticity, and their design would also require the involvement of experts in the field. Teacher assessment has to be systematically complemented by self as well as peer and group assessment. Finally, LSP assessment needs to reach a balance between the evaluation of both process and product as authors highlight.

KEYWORDS: LSP, validity, assessment, self-assessment.

RESUMEN: Hoy en día, las pruebas de lengua para fines específicos están viendo incrementada su demanda, especialmente en lo que concierne al inglés para los negocios. Estas pruebas comprenden tanto los fines académicos como los profesionales u ocupacionales. Este artículo muestra un repaso de los aspectos más importantes que los profesionales deben tener en cuenta al preparar las pruebas, ya que son diferentes a los de las pruebas de lengua generales. Las pruebas de lengua para fines específicos deberían tener la validez necesaria para medir las necesidades profesionales de los estudiantes en términos de autenticidad, y su diseño debería requerir la participación de expertos en el tema. La evaluación del profesor debe ser sistemáticamente completada por la autoevaluación y la evaluación en parejas y en grupos. Para finalizar, la evaluación de pruebas de lengua para fines específicos necesita alcanzar un equilibrio entre evaluación del proceso y del producto, tal y como ponen de relieve los autores.

PALABRAS CLAVE: pruebas de lengua para fines específicos, validez, evaluación, autoevaluación.

ALDERSON AND BANERJEE (2001: 222) trace back the development of specific purpose language testing to the Temporary Registration Assessment Board introduced by the British General Medical Council in 1976 and the English Language Testing Development Unit scales. After an initial stage of theory deficit in LSP testing, nowadays there is a growing demand for LSP tests, especially business English (Thoma, 2010: 77), to the point that Ghent University even launched a project for testing Languages for Specific Purposes, and the European Centre for Modern Languages of the Council of Europe developed the 2008-2011 project GULT, guidelines for university testing in the area of LSP as well as task-based tertiary testing within the framework of the CEFR. LSP tests comprise tests for academic purposes and for occupational or professional purposes. Thoma goes on to refer to the two rationales theoretically justifying the existence of differentiated LSP testing from general purpose language testing: a system-linguistic rationale (LSP and general purpose tests differ in lexical, semantic, syntactic and phonological characteristics, as genre and corpus-based studies on language in specific contexts have supported) and an educational-cognitive rationale (different language abilities for language performance in a particular context). Basturkmen and Elder (2006: 688) point at two practical arguments for LSP testing: «LSP tests make the goals, methods, and outcomes of assessment more transparent and hence more convincing to end users», and «LSP tests are more likely to generate teaching activity which is seen by learners as relevant to their needs». This means that the backwash (the effect on the learning process) achieved by a finely-tuned LSP test design will be beneficial.

Nevertheless, it is not always easy to establish a clear distinction between general purpose and specific purpose language tests. Douglas (2002: 1) highlights the fact that «tests are not either general purpose or specific purpose –all tests are developed for some purpose– but that there is a continuum of specificity from very general to very specific, and that a given test may fall at any point of the continuum», with a growing tendency to move to the more general end. He relies on the two main aspects to distinguish LSP testing from general purpose testing: authenticity of task, and «interaction between language knowledge and specific purpose content knowledge» (*ibidem*: 2). Following the continuum notion of Douglas, O'Sullivan (2002: 3) views all tests as in some ways 'specific'. Tratnik (2008: 5) sees other differences: unlike general English tests, «LSP tests are more likely to be used with adults […] or learners who have already acquired basic knowledge of the language system», and precision in language becomes an essential feature of LSP testing in the particular discipline of the specific purpose. Alderson and Banerjee (2001: 223) find debatable whether LSP tests «are more informative than a general purpose test». In any case, Cumming (2001: 222) points out that «distinctions between specific-purpose and general-purpose approaches to assessment» have emerged as «a salient trend across contexts and program types internationally».

Mislevy and Yin (2009: 263) state that «a LSP test requires capability to use language in a specified space of contexts», which involves «knowledge of substance, practices, conventions, and modes of interaction in those contexts». Davies (2001) does not consider Specific Purposes as registers alone; he characterizes them by their communicative nature and therefore we are in the realm of discourse. He concludes that test usefulness is the only pragmatic criterion on which it is possible to evaluate LSP tests, and questions the evidence of operational tests such as ELTS and IELTS.

Criterion-Referenced testing is very useful for LSP since it is possible to establish the level of ability or domain of content necessary for the target language use (TLU) situation, and assess the achievement of this level by a student (Lynch, 2003: 32), regardless the group he may belong to (the traditional Norm-Referenced Measurement). Douglas (2001: 173) believes that LSP assessment criteria should be derived from an analysis of the TLU situation (the pragmatic concept of 'language in use') and the concept of 'indigenous assessment criteria' («those used by subject specialists in assessing the communicative performances of apprentices in academic and vocational fields», which can be investigated and described by means of grounded ethnography in based context research), beyond a mere theoretical understanding of communicative language ability: «LSP test-developers can and do find out in detail during the test development process what situations the test-takers will find themselves in and are able to draw on the linguistic and situational features» the students will need to be proficient in. Alderson and Banerjee (2001: 223) caution that the need for in-depth analysis of the TLU situation make LSP tests «time-consuming and expensive to produce». Test tasks should be the translation of the tasks that language users typically perform in the target situation, the ones dealt with in the class (Gnutzmann, 2009: 530). Ellis has noted (2004: 292) that the consideration of the psycholinguistic dimensions of tasks and how they affect performance is missing from this approach, although «the way language is stored affects the way it is used in real-time performance». Needs analysis is a key tool to inform LSP course and test design (Ingham and Thighe, 2006: 6), and to indicate both pedagogically and assessment oriented tasks. Design patterns differ from test specifications by building assessment tasks to address situations of language use for different purposes or circumstances: «a unique feature of a test of English for Specific Purposes (ESP) is that the tasks must stimulate the interaction between the examinee's language ability and specified content knowledge» (Wei, Mislevy and Kanal, 2008: 12).

The fundamental problem of ESP testing, according to West and Tompos (2000: 195), is «that of designing tests which assess the specific branches of English […] while at the same time ensuring that the different tests are comparable in coverage and language level». The starting point of ESP testing is validity: if the test is valid to measure the learner's professional needs, with authenticity (both situational and interactional) as the cornerstone for its content validity. Although Thoma (2010: 331) warns: «To date, there is no LSP test validation study that used a stardadized test of background knowledge as a control». Effective LSP test design would also require the involvement of experts in the field (Gnutzmann, 2009: 530), «subject specialist informant procedures» (Douglas, 2001: 182), and teacher assessment has to be systematically complemented by the student's self-assessment as well as peer and group assessment, although Januleviciene and Kavaliauskiene (2007: 10) found a mismatch between «an ESP novice's self-assessment of professional language proficiency and actual performance in real-life job-related situations», thus concluding the importance of training ESP students to assess themselves in order to raise autonomy and intrinsic motivation and engage students in their own learning process. In team work, a grade can be given both to individual students and to the group as a whole. Besides, domain analysis is an important content of test design in LSP assessment (Fulcher and Davidson, 2007: 371), thus pursuing a balance between «the assessment of achievement (school-oriented, learning-oriented) and the

assessment of proficiency (real world-oriented, outcome-oriented)» (Council of Europe, 2001: 192), in the same way as it is necessary to reach a balance between the assessment of both process and product, although assessment practices in education have traditionally been product oriented. Donna (2000: 113) suggests «including a rate for body language in any assessment», e.g. when doing presentations, given the importance of nonverbal communication and kinesics in ESP.

Weigle (2005: 80) summarizes the three options test designers have for construct definition as regards topical knowledge: (a) exclude it from the construct definition, as it is the case when test takers are not expected to have similar knowledge and decisions are based only on language ability; (b) include it as part of the construct definition, as in LSP programmes, where test takers are expected to have similar topical knowledge; (c) separate topical knowledge and language ability, as in content-based language courses.

In LSP courses, the previous subject knowledge that learners bring to the class allows them to «cope with texts well above their estimated proficiency level» (Mishan, 2005: 63), and while this must be exploited by the teacher as a useful help to the learning process, it must also be weighed whether it could be a distorting element when testing. Alderson and Banerjee (2001: 222) see a crucial difference between general purpose and specific purpose testing in the fact that, in the former, background knowledge is «a confounding variable that contributes construct-irrelevant variance to the test score», whereas in the latter it is essential. Douglas (2001: 46) sees background knowledge, as «a necessary, integral part of the concept of "specific purpose language ability"», since it forms part of the construct in contexts of an academic, professional or vocational kind, the same as strategic competence, which is the link between the situational context and the language and content knowledge necessary for communication. Some studies have observed that the effect of background knowledge varies according to the level of language proficiency, or linguistic thresholds (Krekeler, 2006), especially for intermediate candidates (inseparability of content and language knowledge).

In the particular case of EAP, Brindley and Ross (2001: 149) distinguish two kinds of assessment: achievement («to determine the extent to which learners have learnt what has been taught during a course of instruction») and proficiency («to establish the extent to which learners can use the language for their intended purposes»). Assessment is equally important for evaluating individual attainment as well as for programme evaluation (external benchmarks for assessing EAP programme 'gains').

Portfolios or web-based e-portfolios also have many potential applications in LSP assessment (Douglas, 2002: 241), including the meta-cognitive strategy of reflective practice and self-assessment involved in the selection of learner work. Advocates of portfolio assessment remark the collection of a wide range of texts that it allows for. It also provides innovation in (self-) assessment practices related to autonomous and lifelong learning (Durán and Cuadrado, 2007: 121), promoting their self-confidence and improving critical thinking skills. Hismanoglu and Hismanoglu (2010: 682) state that the ELP can stimulate ESP learners «to take their own language learning initiatives as well as enabling them to carry on their foreign language learning and also foreign use beyond the borders of the classroom». Donna (2000: 301-302) designs a feedback sheet for portfolio assessment, and suggests possible samples of the student's work to include in a business English portfolio, like e-mails, faxes, memos and reports. She

also enumerates several advantages to using portfolio assessment, such as: tasks can be done in class, for homework or in the workplace, and at the student's own pace, and this can accentuate the results-oriented nature of the ESP course.

Benesch (2001: x) reminds us that the critical theory and critical approaches to ESP pedagogy have raised questions of rights analysis, that is, the issue of power that «relates to who makes the decisions about the content of the course and the nature of teaching and assessment»; students «can and should help shape the nature of the course and the forms of assessment».

According to Thoma (2010: 77), the TOEIC, launched by ETS in 1979, was the first standardized language test for business, consisting in multiple choice items within a psychometric-structuralist framework. However, other important business language tests like the BEC and BULATS follow the pattern of communicative language testing. Douglas (2002: 1) considers LSP testing a case of communicative testing, as it is «based on a theoretical construct of contextualized communicative language ability». In the 80s, the ELTS focused on real-life language use and consequently on field-oriented tests with an ESP approach (Davies, 2007: 79). Tratnik (2008: 5) defines any ESP test as «a performance test assessing the skills needed». The examinee also has to show his/her pragmatic and sociocultural knowledge to adjust the language to a specific situation or audience (Wei, Mislevy and Kanal, 2008: 2). At present, computerized testing by means of internet-based ESP testing platforms like iBT TOEFL is very common and allows for assessing thousands of students internationally, and the tendency is bound to grow thanks to the advent of Learning Management Systems like Moodle. Computer-adaptive testing (CAT) can prove useful for LSP learners, as it adapts the items according to the ability of the test taker.

Guzmán and Alberola (2001: 255) propose an assessment process in an ESP setting that starts at the beginning of the course with a placement test to divide students in groups according to their level of proficiency; goes on with continuous assessment adjusted to each group; and ends with a global test. All testing is «communicative, skills-based, contextualized and based on real life situations». Piqué-Angordans and Posteguillo (2006) focus their attention in peer evaluation in medical English academic genres, where positive and negative assessment is conveyed by means of linguistic devices such as hedging or reporting verbs. Ekbatani (2011: 87) sees paired presentations as an especially helpful method for assessing oral proficiency in ESP courses. Douglas (2005: 857) argues for the use of extended input from genuine sources in specific purpose tests, as well as the use of trained interlocutors in role-plays, since simulations lack pragmatic reality. In the case of simulation and gaming methodology aimed at business communication, Garcia-Carbonell and Watts (2009: 304) consider necessary to establish, following the principle of optionality underlying the negotiated syllabus, who will assess, with what criteria and at what moment, before the beginning of an activity which is going to be assessed: «Assessment can be very detailed with analytical rubrics or it can be done holistically by impression or by a half-way compromise using descriptors to judge performance. The criteria can and should be negotiated among those involved».

As it has happened extensively in the past, it is realistic to expect that ESP assessment and testing will undergo important changes in the next few years, as a logical response to the changes in ESP methodology itself.

References

ALDERSON, J. Charles and BANERJEE, J. (2001) Language testing and assessment (part I). *Language Teaching*, 34 (4), 213-236.

BASTURKMEN, H. and ELDER, C. (2006) The practice of LSP. In Alan DAVIES and Catherine ELDER (eds.) *The handbook of applied linguistics* (pp. 672-694). Oxford: Blackwell.

BENESCH, S. (2001) *Critical English for academic purposes*. New Jersey: Lawrence Erlbaum.

BRINDLEY, G. and ROSS, S. (2001) EAP assessment: Issues, models, and outcomes. In John FLOWERDEW and Matthew PEACOCK (eds.) *Research perspectives on English for Academic Purposes* (pp. 148-167). Cambridge: Cambridge University Press.

COUNCIL OF EUROPE (2001) *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

CUMMING, A. (2001) ESL/EFL instructors' practices for writing assessment: specific purposes or general purposes? *Language testing*, 18 (2), 207-224.

DAVIES, A. (2001) The logic of testing languages for specific purposes. *Language testing*, 18 (2), 133-147.

— (2007) Assessing academic English language proficiency: 40+ years of U.K. language tests. In Janna D. FOX, Mari WESCHE, Doreen BAYLISS, Liying CHENG, Carolyn E. TURNER and Christine DOE (eds.) *Language testing reconsidered* (pp. 73-88). Ottawa: University of Ottawa Press.

DONNA, S. (2000) *Teach business English*. Cambridge: Cambridge University Press.

DOUGLAS, D. (2001) Language for Specific Purposes assessment criteria: where do they come from? *Language Testing*, 18 (2), 171-185.

— (2001) Three problems in testing language for specific purposes: Authenticity, specificity and inseparability. In Catherine ELDER and Alan DAVIES (eds.) *Experimenting with uncertainty* (pp. 45-52). Cambridge: Cambridge University Press.

— (2002) *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.

— (2005) Testing languages for specific purposes. In Eli HINKEL (eds.) *Handbook of research in second language teaching and learning* (pp. 857-888). New York: Routledge.

DURÁN ESCRIBANO, P. and CUADRADO ESCLÁPEZ, G. (2007) Developing language descriptors to match the writing competence of engineering students. In Ana BOCANEGRA VALLE, María del Carmen LARIO DE OÑATE and Elena LÓPEZ TORRES (eds.) *English for Specific Purposes: Studies for classroom development and implementation*, (pp. 105-124). Cádiz: Universidad de Cádiz.

EKBATANI, G. (2011) *Measurement and evaluation in post-secondary ESL*. New York: Routledge.

ELLIS, R. (2004) *Task-based language learning and teaching*. Oxford: Oxford University Press.

FULCHER, G. and DAVIDSON, F. (2007) *Language testing and assessment*. New York: Routledge.

GARCÍA-CARBONELL, A. and FRANCES, W. (2009) Simulation and gaming methodology in language acquisition. In Victoria GUILLÉN-NIETO, Carmen MARIMÓN LLORCA and Chelo VARGAS-SIERRA (eds.) *Intercultural business communication and simulation and gaming methodology* (pp. 285-316). Bern: Peter Lang.

GNUTZMANN, C. (2009) Language for specific purposes vs. general language. In Karlfried KNAPP and Barbara SEIDLHOFER (eds.) *Handbook of foreign language communication and learning* (pp. 517-544). Berlin: Walter de Gruyter.

GUZMÁN GIL, E. and ALBEROLA COLOMAR, P. (2001) An assessment system proposal for an English for tourism course. In S. POSTEGUILLO, I. FORTANET and J. C. PALMER (eds.) *Methodology and new technologies in languages for specific purposes*, (pp. 251-260). Castellón: Universitat Jaume I.

HISMANOGLU, M. and HISMANOGLU, S. (2010) The European Language Portfolio in ESP classes: A case study of learner reflection and self-assessment. *European Journal of Social Sciences,* 12 (4), 671-684.

INGHAM, K. and THIGHE, D. (2006) Issues with developing a test in LSP: The International Certificate in Financial English. *Research Notes*, 25 (2), 5-9.

JANULEVICIENE, V. and KAVALIAUSKIENE, G. (2007) Self-assessment of vocabulary and relevant language skills for evaluation purposes. *Filologija*, 15 (4), 10-15.

KREKELER, C. (2006) Language for special academic purposes (LSAP) testing: The effect of background knowledge revisited. *Language testing*, 23 (1), 99-130.

LYNCH, B. K. (2003) *Language assessment and programme evaluation.* Edinburgh: Edinburgh University Press.

MISHAN, F. (2005) *Designing authenticity into language learning materials.* Bristol: Intellect Books.

MISLEVY, R. J. and YIN, Ch. (2009) If language is a complex adaptive system, what is language assessment? In Nick C. ELLIS and Diane LARSEN-FREEMAN (eds.) *Language as a complex adaptive system* (pp. 249-267). Chichester: Wiley.

O'SULLIVAN, B. (2002) Some theoretical perspectives on testing language for business. *Research Notes*, 8 (01), 2-4.

PIQUÉ-ANGORDANS, J. and POSTEGUILLO, S. (2006) Peer positive and negative assessment in medical English written genres. In Maurizio GOTTI and Françoise SALAGER-MEYER (eds.) *Advances in medical discourse analysis: oral and written contexts* (pp. 383-406). Berlin: Peter Lang.

THOMA, D. (2010) *Strategic attention in language testing: Metacognition in a yes/no business English vocabulary test.* Frankfurt: Peter Lang.

TRATNIK, A. (2008) Key issues in testing English for specific purposes. *Scripta Manent*, 4 (1), 3-13.

WEI, H.; MISLEVY, R. J. and KANAL, D. (2008) An introduction to design patterns in language assessment. *PADI Technical Report*, 18.

WEIGLE, S. C. (2005) *Assessing writing.* Cambridge: Cambridge University Press.

WEST, R. and TOMPOS, A. (2000) From generic to specific: a genre-based approach to ESP testing. In Mike BEAUMONT and Teresa O'BRIEN (eds.) *Collaborative research in second language education* (pp. 195-206). Stoke on Trent, UK: Trentham.