IS ESP STANDARDIZED ASSESSMENT FEASIBLE?

Jolita Šliogerienė

Vilnius Gediminas Technical University, Lithuania
E-Mail: jolita.sliogeriene@vgtu.lt

Abstract. *This paper provides an overview of assessment practice as it relates to English for Specific Purposes (ESP), and which is used in a variety of higher education settings. The notion of benchmarking in ESP standardized assessment is discussed, and assessment standards leading to quality assurance are described. The challenge for educators is to agree on a common assessment framework in view of the ongoing debate on ESP benchmarking and unified assessment criteria (nationally or internationally), which is also compared to CEFR. The author uses case study analysis to focus on student assessment policy and practice in Alberta, Canada, as well as other selected countries. It is significant that, today, a number of ESP assessment models are based only partially on the main foreign language assessment principles. Accordingly, this paper provides an overview of such principles, their descriptors and best practice in ESP assessment. The main aim of the research is therefore to understand current assessment practices as well as to develop a standardized benchmarking for ESP teachers. The paper proposes a model of ESP standardized assessment based on the studied reference, ESP practices used in different countries, as well as standards of assessment in general.*

Key words: *ESP, standardized assessment, benchmarking, formative and summative assessment.*

1. INTRODUCTION

New challenges continue to confront higher education policy, and these require different approaches to the traditional assessment methods used in the field of English for Specific Purposes (ESP). In recent years, there has been much discussion about which assessment method should prevail at higher education institutions: summative or formative. One issue with which few scholars have engaged is whether or not ESP assessment at the national and/or international level should be benchmarked and standardized. For higher education assessment the debate has taken on renewed importance due to the increasing internationalization of education, which continues to challenge the competence of teachers. Wenger (2001), who is also cited in (Irja Leppisaaria, Leena Vainiob et al., 2011), asserts that new technologies have extended the reach of our interactions beyond the geographical limitations of traditional communities. With the shift to virtual learning and assessment platforms, and the impact of social media on educational processes, there is a growing demand for benchmarking, especially if viewed from the perspective of the Bologna process and European Higher Education Area (EHEA) objectives for 2020. These establish priorities such as a professional and

employability approach, as well as a social and mobility dimension (Leuven and Louvain-la-Neuve Communiqué, 2009).

What, then, is benchmarking? The term can best be understood as a process in which an organization (educational institution) compares its outcome-based results and achievements against other institutions. The intention is to establish best practice and thereby improve performance. According to Irja Leppisaaria, Leena Vainiob et al., benchmarking supports learning opportunities garnered from personal experience and from the experience of others, which is something that Malderez and Wedell (2007) consider crucial in the teacher's ongoing learning. With this in mind, the benchmarking process becomes a learning space, created through communication and the exchange of ideas, knowledge, experiences and emotions in a reflective and authentic way (Boud, 2006; Docherty, Boud & Cressey, 2006).

The subject of benchmarking or standardized testing points is integral to the *idea* of quality management in education. There are many questions and points of interface to consider, as well as a plethora of tools, techniques and systems that organizations can use to introduce and develop an *idea*. If the native speaker is taken as a benchmark by which to assess L2 performance, one challenge is a consideration of language variations. Standardization of any language is commonly viewed as an idealization (Milroy & Milroy, 1991), with variation accepted as an inevitable reality. It is important to note that this paper does not deal with the question of whether increased awareness of the existence of differences between mono-competent and multi-competent systems is needed, nor acknowledgement of the fact that our own linguistic intuitions are constrained not only by our regional, socioeconomic, and gendered identities, but also by our own language learning experiences. This acknowledgement often results in the interest of language variation in professional development and different teachers' competences in the field of ESP standardization. The paper explores the problem of ESP standardized assessment and assessment standards leading to quality assurance. *The main aim of the research is to determine the current situation in ESP assessment used by ESP teachers in order to define whether standardized ESP assessment could be established.* The main object of the research is therefore ESP assessment.

2. LINKING FORMATIVE AND SUMMATIVE FUNCTIONS IN THE DESIGN OF ESP ASSESSMENT BENCHMARKING

It is apparent that higher education institutions place little emphasis on a consideration of concepts such as: Assessment of Learning (AoL); Assessment for Learning (AfL); or Assessment as Learning (AaL). Whether at national or international level, educators have found it difficult to establish a common framework for assessment for ESP in terms of benchmarking and unified assessment criteria. All institutions follow CEFR (Common European Framework of Reference) to assess general English, yet what criteria and techniques should they use for ESP assessment? Is it simply a matter for each higher education institution to decide their own?

The paper will focus on the formative assessment policy and practice *through an assessment for learning (AfL)* as well as on the summative assessment challenges *through an assessment of learning (AoL).* According to Sin Wang Chong (2017) AoL provides reliable and valid ways to measure, summarize, and evaluate students' acquired skills and

knowledge, as opposed to how assessment can be implemented to improve learning and instruction. Teachers in higher education are expected to analyse the final set examination with the intended learning outcomes as prescribed in the course outline. The quality of this summative assessment is not measured on how accurately it reflects student learning in a given period of time, but on how faithfully it adheres to the established examination system, social expectations, and expectations of the school administrators (Sin Wang Chong, 2017). The primary role of formative assessment is to help students understand the progress they are making during a specific period of time. A variety of measurement instruments can be used to evaluate learners' progress in order to help them project their future studies and outcomes to be achieved – feedback in different forms, portfolio, reflection pages, learning journals, logs, blogs, and all social media platforms. Students play an important role in AfL by reflecting on their own learning, although Carless (2007) has argued that the role of the teacher is of primary importance in the effective implementation of AfL, simply because teachers are the mediators "in enhancing student learning. Thus, improvements in the implementation of formative assessment depend largely on a teacher's understandings of principles and practice in formative assessment." (p. 172) as cited in Sin Wang Chong, 2017, p.5.

Following some case study analysis with a focus on student assessment policy and practice in Alberta, Canada (J. Brandon and M. Quarin Wright, 2012), there were some major goals set before the establishment of standards in assessment for learning. The authors present student assessment policy and practice within the context of the 2009 Alberta Student Assessment Study (ASAS) (Webber et al. 2009). The main goals were related to curricular learning outcomes and performance standards as well as reporting of levels of achievement within grade, the questions how external tests and classroom-based assessment of student achievement can be optimally used to inform decisions regarding student programme needs and how they relate to the improvement of learning opportunities for students. The involvement of educational leadership in order to facilitate more effective classroom assessment as well as accurate and meaningful reporting of assessment information has to be considered. The need for professional development models and an answer to the question: *Where do we go from here?,* were also observed ((J. Brandon and M. Quarin Wright, 2012). The authors designed a conceptual framework depicting the relationships among educational research, policy and practice and a pyramidal structure in which policy, leadership and teaching influence student learning. Having a certain degree of teaching and leadership practice as well as educational policy, a theoretical model for assessment policy could be designed (Fig.1). The figure illustrates that the core of the assessment policy and learning practice is research-informed assessment standards. Practice and policy makers in Lithuania recently commenced discussions in this area, that there should be valid and reliable research-based results, which are applicable in every educational institution. The elements presented in the chart demonstrate that teaching with all forms of assessment, as well as content and process, should be unified.
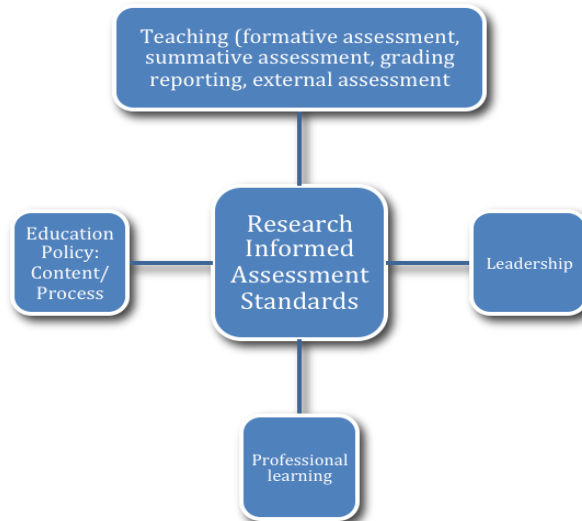
Fig. 1 Assessment for policy and practice learning

3. ESP ASSESSMENT STANDARDS IN QUALITY ASSURANCE FRAMEWORKS

The arrival of a multidisciplinary approach, the core of which was constructivism, signaled a shift in language teaching and learning. This can be understood within the context of certain schools of thought, notably structural linguistics and behavioral psychology, which focused mainly on description, empiricism, surface structure, conditioning and reinforcement. Subsequently, this progressed to generative linguistics and cognitive psychology, with an emphasis on acquisition, inter-language, universal grammar, competence and deep structure. This approached reached the height of its popularity in the 1980s. Yet, constructivism and its associated ideas are nothing new; Jean Piaget (1970) and Lev Vygotsky (1978) both embraced such ideas. Regardless of social constructivism and the zone of proximal development (ZPD), there was a shift in thinking to new themes and concepts. These included interactive discourse, sociocultural variables, cooperative learning, and construction of meaning, to name just a few (H. Douglas Brown, 2007). In due course, these ideas were further refined and goals identified with a focus on the perspectives of educational assessment as enabling, whereby the learner is central and assessment encounters a lot forms. This effected a major paradigm shift in the movement towards a student-centered approach and especially the red-design of assessments in order to support quality learning.

Of continuing interest to those involved in higher education is whether or not a quality assurance framework is actually needed, and if so, which model should be adopted. A number of factors contributed to changes in quality assurance processes. According to M. Shah and L. Jarzabkowskib (2013), these include: (1) the introduction of various policy instruments by governments to increase university productivity with declined funding; (2) increased participation of students in higher education; (3) the emergence and growth of non-university providers; (4) increased demand and use of ICT in learning; (5) increased

student diversity with different expectations and experience; (6) internationalization of higher education with increased student mobility, offshore campuses, and partnership delivery; and (7) diversion of government funding in areas of strategic priority due to unpredicted global economy and the uncertainty surrounding future growth (Mahsood Shah and Lucy Jarzabkowskib, 2013). A case in point is Lithuania where the rising cost of tuition, skyrocketing student debt, and falling enrollments, have placed ever greater demands on institutions of higher education. For the latter, the challenges are the same, especially at a time when the quality assurance process has only just commenced. Whilst a system of programme accreditation has been implemented, the quality in every area of study is under discussion though descriptors for separate study fields have also been identified. Every descriptor of a study field has foreseen a requirement, e.g. "Ability to communicate with engineering or any other field community and the general public in *standard Lithuanian and at least one foreign language*". Here we come across a great variety of interpretations and assessment templates with no precise national benchmarking, let alone international unification as far as ESP assessment is concerned. Before designing a certain framework for ESP assessment, a number of unified criteria, assessment principles should be established and the validity of accountability models should be discussed.

The importance of assessment in a curriculum context should not be underestimated. Moreover, the value of experts should be widely acknowledged in any educational context (J. Gardner, 2012). Following the establishment of theoretical frameworks and practice across different higher education institutions, we can observe the emergence of a number of ESP assessment models that are only partially based on the main foreign language assessment principle, which can be expressed as follows:

Table 1 ESP assessment principles

Principle	Descriptor	ESP assessment state-of the-art
Assessment should be valid.	Validity should assure that all tasks and associated criteria effectively measure student attainment of the intended learning outcomes at the appropriate level.	Recording of ESP competence and its alignment with study programme outcomes and achievements in a given subject is usually done as mechanical input into tables. Usually, neither teachers nor students are active participants in the process.
Assessment should be reliable and consistent.	There is a need for assessment to be reliable, but this requires clear and consistent processes for the setting, marking, grading and moderation of assignments.	Since a new teacher would usually apply their own assessment criteria, data can be unreliable. This may be due to ESP being taught in small subgroups where assessment varies between each class.
Information about assessment should be explicit, accessible and transparent.	Clear, accurate, consistent and timely information on assessment tasks and procedures should be made available to students, staff and	Generally, ESP teachers assess their own students; external assessors or examiners are not available during exams.

	other external assessors or examiners.	
Assessment should be inclusive and equitable.	As far as possible, and without compromising academic standards, inclusive and equitable assessment should ensure that tasks and procedures do not disadvantage a group or individual.	Equity is almost guaranteed in ESP assessment insofar as race and gender are concerned and procedures do not (or should not) disadvantage any student.
Assessment should be an integral part of programme design and should relate directly to the programme aims and learning outcomes.	Assessment tasks should primarily reflect the nature of the discipline or subject, but should also ensure that students have the opportunity to develop a range of generic skills and capabilities.	A range of generic competences developed during the ESP course is also demonstrated, however a range of very specific language (ESP) at a higher level (C1) may not be properly reflected for a variety of reasons, including: the exam is usually in the first year of studies when subject-based knowledge is in the introductory stage; and students demonstrate General English competence with some touches of ESP, etc.
The amount of assessed work should be manageable.	The scheduling of assignments and the amount of assessed work required should provide a reliable and valid profile of achievement without overloading staff or students.	The principle of efficiency helps to identify an important problem, namely a different understanding of how learning outcomes are achieved, contact hours and number of credits devoted to ESP learning from three parties: study programme designers, ESP teachers and students. This principle is also differently interpreted at the inter-institutional level.
Formative and summative assessment should be included in each programme.	Formative and summative assessment should be incorporated into programmes to ensure that assessment purposes are adequately addressed.	Different institutions apply various formative assessment method, while summative assessment is often treated as attaining a certain level of ESP acquisition.
Timely feedback that promotes learning and facilitates improvement should be an integral part of the assessment process.	Students should receive advance notice of the nature, extent and timing of feedback for each assessment task.	Students often receive feedback on submitted formative assessment tasks, and on summative tasks in different forms, depending on the institution. Feedback on formative assessment needs to be radically improved, as well as final grading.

Staff development policy and strategy should include assessment.	Everyone involved in the assessment must be competent to undertake their roles and responsibilities.	ESP assessors and examiners are well acquainted with CEFR, and know their roles, yet the majority have received inadequate training in what the valid and reliable ESP assessment should be, as well as what instruments to use to follow benchmarking and quality-based standardized ESP assessment.

4. ESP ASSESSMENT: STANDARDIZED OR INSTITUTION-BASED AMATEUR GRADING?

English for Specific Purposes assessment must answer some basic questions, namely: why is the assessment performed, and what is going to be assessed? For J. Brandon and M. Quarin-Wright, every assessment standard must exhibit a clear picture of evidence-based policy or practice that will be used in a particular assessment. Note that assessment standards discussed in the paper are adapted using the model of student assessment policy and practice in Alberta (J. Brandon and M. Quarin-Wright, 2012).

Quality teaching as situated, collective expertise in action

Formative assessment as generative and informative teaching

Summative assessment, grading and reporting as consistent, accurate, and outcome-referenced descriptions of learning

External assessment as complementary outcome- referenced descriptions of learning;

Principal quality practice guideline

Quality education policy content

Quality education policy process

Fig. 2 The research – informed assessment standards

The first standard (Fig. 2) focuses on three main criteria: contingency of the dynamic interplay of content, teacher, learner and context; involvement of professional commitment to collegial practice and reflection over time; accesses to scientific and artistic pedagogic tools in a fluent, seamless, holistic ad constantly evolving pattern of practice.

The second standard – *formative assessment* as a key aspect of teaching and learning that provides students with clear pictures of progress and how to improve during the learning process.

The third standard - *summative assessment*, grading, and reporting that provide students and social partners, employers with clear pictures of achievement in relation to learning outcomes in the programme of at the end of a learning episode: provides consistent, accurate and outcome-referenced descriptions of learning; based upon informed professional judgment using varied assessment tools to show best available evidence of learning; fosters student involvement in, reflection on and ownership of the learning process.

The fourth standard - *external assessments* as complementary outcome - referenced descriptions of learning that provide data over time to policymaking and curriculum development; provide data over time to be used in combination with classroom, school and jurisdiction data to inform longer term instructional, school and system improvement planning; one of many assessment tools to inform professional judgment in relation to summative assessment, grading and reporting. The original model had two more standards based on professional learning as coherent, supplementary capacity building which emphasizes staff learning, integration of ongoing opportunities for reflection, professional dialogue, and continuous pedagogical learning

A standard of *Education policy content* is measured by research, exemplary practice and policy learning. Lastly, *Quality education policy process* is developed through informed design, dialogic adoption, and implementation as learning and meaningful outcomes. The original model had two more standards based on professional learning as coherent, supplementary capacity building which emphasizes staff learning, integration of ongoing opportunities for reflection, professional dialogue, and continuous pedagogical learning.
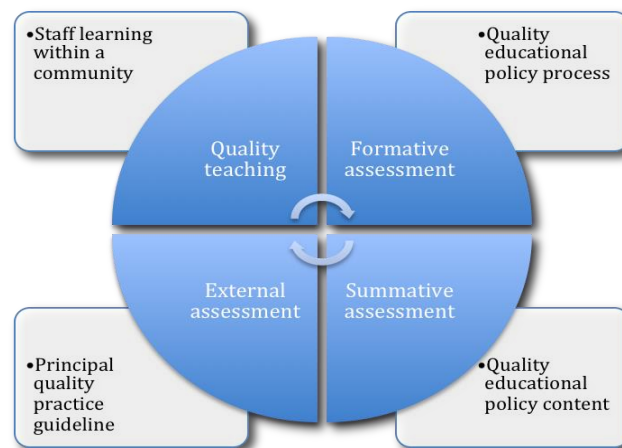

Fig. 3. A model of ESP standardized assessment

Based on the studied reference and having examined different practices used in various countries in the field of ESP assessment as well as standards applied in assessment in general, a model of ESP standardized assessment is proposed (Fig.3).

7. CONCLUSION

Having analysed the theoretical frameworks and practice used in standardized assessment, the following conclusions can be drawn: 1) the importance of assessment in a curriculum context cannot be underestimated and the value of external experts should be widely acknowledged in any educational context; 2) the main goals related to curricular learning outcomes and performance standards and reporting of levels of achievement within grade must be clear from the outset; 3) the questions are: what is the optimal use of both external tests and classroom-based assessment of student achievement if we are to reliably inform the community and/or society if the assessment is valid, reliable, consistent and equivalent to those assessments conducted by institutions that are different.

This paper has argued that if evaluators are to understand appropriate standardized assessment, they should follow the same theoretical frameworks and practice as well as principles of ESP assessment used in different higher education institutions. Standards of ESP assessment described in this paper are critical in understanding what benchmarking and standardized assessments are. A number of standards are described, but only some of them are adapted in the context of ESP standardized assessment. The novelty of this paper lies in the design of the model of ESP standardized assessment with its main components: quality teaching, formative assessment, summative assessment and external assessment. Four subcomponents dealing with the quality assurance in ESP assessment are presented respectively: staff learning within a community, quality educational policy process, quality educational policy content and principal policy practice guideline. Both a separate component and subcomponent are recommended for a more detailed analysis in every country in order to reach a common understanding of standardized ESP assessment.

REFERENCES

Alberta Education (2007). Presentation to the Programme Standards and Accountability Stakeholder Advisory Committee. November 2007, Edmonton, AB.

Boud, D. (2006). Creating a space for reflection. In D. Boud, P. Cressey, & P. Docherty (Eds.), Productive reflection at work (pp. 158–169). London and New York: Routledge.

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. Innovations in Education and Teaching International, 44(1), 57–66. doi:10.1080 /14703290601081332

Davidson, F. & B. K. Lynch (2002). Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications.London: Yale University Press.

Docherty, P., Boud, D., & Cressey, P. (2006). Lessons and issues for practice and development. In D. Boud, P. Cressey, & P. Docherty (Eds.), Productive reflection at work (pp. 193–206). London and New York: Routledge.

H. Douglas Brown. (2007). Principles of Language Learning and Teaching. Pearson Education, Inc. https://www.youtube.com/watch?v=cQWDJeaswhU

International Language Testing Association (ILTA) (2002). http://www.surrey.ac.uk/ ELI/ILTA/faqs

Irja Leppisaari , Leena Vainio , Jan Herrington & Yeonwook Im (2011) International e-benchmarking: flexible peer development of authentic learning principles in higher education, Educational Media International, 48:3, 179-191, DOI: 10.1080/09523987.2011.607321

John Gardner (2012). Assessment and Learning. Quality assessment practice. Sage Publications Ltd. London. p. 103

Leuven and Louvain-la-Neuve Communiqué (2009). *The Bologna Process 2020: The European Higher Education Area in the new decade. Commmuniqué of the Conference of European Ministers Responsible for Higher Education. Belgium.* Retrieved on 05-04-2015 from http://www.ehea. info/Uploads/Declarations/Leuven_Louvain-la-Neuve_Communique_April_2 009.pdf

Mahsood Shah & Lucy Jarzabkowski (2013) The Australian higher education quality assurance framework, Perspectives: Policy and Practice in Higher Education, 17:3, 96-106

Milroy, J., &Milroy, L. (1991). Authority in language: Investigating language prescription and standardization. London, UK: Routledge.

Piaget, J. (1970). The science of education and the psychology of the child. New York: Basic Books

Robert Wood (1993). Assessment and testing. Cambridge University Press. P.273

Sin Wang Chong (2017): Three Paradigms of Classroom Assessment: Implications for Written Feedback Research, Language Assessment Quarterly, DOI: 10.1080/15434303.2017.1405423

Vygotsky, L. (1978). Mind in society. The development of higher psychological processes. Cambridge, M.A:Harvard University Press

Wenger, E. (2001). Supporting communities of practice. A survey of community-oriented technologies. Retrieved 2 July 2018 from http://www.ewenger.com/tech.