ASSESSMENT OF ACADEMIC STAFF PERFORMANCE IN ENGLISH MEDIUM INSTRUCTION

VITA KALNBĒRZIŅA and INDRA KARAPETJANA
University of Latvia, Latvia

Abstract. One of the internationalisation strategies for higher education institutions is internationalisation of their curricula by delivering English-medium instruction programmes. These internationalization efforts can be successful if support for the language needs of all stakeholders involved is provided. English language proficiency assessment of academic staff is an essential prerequisite to the implementation of a high-quality study process in the English language. Therefore, the aim of this paper is to validate the assessment system of academic staff performance in English medium instruction at a university in Latvia and the use of mediation strategies as a basis for the assessment system development. The qualitative and quantitative analysis of the test results suggests that the integrated assessment tasks demonstrate higher internal consistency and higher correlation with the test results and the performance of the academic staff in the international setting.

Key words: English-medium instruction, mediation strategies, university, academic staff, English performance assessment

INTRODUCTION

Higher education (HE) strategies in the European Union (EU) are affected by worldwide globalisation and internationalisation of HE. Internationalization, as ‘the integration or intercultural dimension into the tripartite mission of teaching, research and service functions of Higher Education’ (Maringe and Foskett 2010: 1), is promoted at international, national and institutional levels.

Internationalisation can be perceived as recruiting and exchange of international students to pursue their undergraduate and postgraduate degree programmes, as attracting and employing international academic staff and as preparing home students for an international career. In order to facilitate international competitiveness, internationalisation strategies for higher education institutions (HEIs) in Latvia are formulated in the policy document “The National

English-medium instruction (EMI) programmes, which have been defined as the teaching of academic subjects using the medium of the English language in countries where English is not the national language, have been seen as a mechanism for internationalising the educational offer and increasing international mobility. There has been an increasing tendency to introduce English-medium instruction programmes in the world. For instance, universities in Finland, Sweden and the Netherlands have the largest proportion of EMI programmes (Wächter and Maiworm, 2014). The University of Latvia also offers bachelor study programmes, master study programmes, professional study programmes, and doctoral study programmes in English.

EMI programmes require a high English language proficiency level not only on the part of students but also academic staff, whatever their mother tongue is. Since most academic staff who teach through EMI are not native speakers of English, they often lack adequate English language skills. Therefore, teaching academic subjects in English has become a serious academic quality issue and must be addressed by universities.

Recently, many higher education institutions in Latvia, including the University of Latvia, have signed the ‘Agreement between the Ministry of Education and Science of the Republic of Latvia and Higher Education Institution Good Practice of Attracting International Students and Delivering Studies’ (Ministry of Education and Science of the Republic of Latvia, n.d.). To ensure the delivery of a high-quality study process, this document stipulates that a HE ‘institution shall ensure that study programmes for international students are implemented by qualified teachers whose level of foreign language in relation to implementing the study programme is high, preferably level C1’ (ibid.). Thus, English language skills of academic staff are an important pre-requisite for a successful implementation of EMI programmes (cf. Kalnbērziņa, 2017).

There have been many studies that analyse EMI implementation in various countries (e.g. Dearden, 2014; Earls, 2016; Fenton-Smith, Humphreys and Walkinshaw, 2017; Margic and Vodopija-Krstanovic, 2017). However, only a few researchers (e.g. Carrió-Pastor, 2020) have attempted to focus on assessing English language proficiency of academic staff who are already teaching or who are expected to teach their subjects in English at some point in their academic careers.

Therefore, in this article, we analyse the construct and the results of the EMI test of academic language skills developed for the academic staff of the UL. This paper does not aim at measuring whether the academic staff have gained
proficiency by taking English language classes. Whether or not there has been sufficient training for academic staff at C1 or B2 levels to be able to teach their academic subjects in English is open to debate and is not the focus of this paper. Instead, we focus on the English language performance test for academic staff, its construct and its validation methods.

ASSESSMENT SYSTEMS FOR ENGLISH MEDIUM INSTRUCTION

There are a number of internationally recognized English language tests, but they are general English tests that do not measure the capacity of academic staff to deal with the English language in an academic context, that is, when using EMI, researching and publishing research articles. For this reason, universities across the world have been working on the development of curricula and English language tests for academic staff (see Table 1 below) to ensure that they have a good command of language so that their students can receive high quality teaching. The approaches adopted by universities differ in their contents: from testing oral proficiency in Denmark (Test of Oral English Proficiency of Academic Staff (TOEPAS)) to the focus on writing research papers in Australia (Scholarship of Teaching and Learning (SoTL) ‘movement’ to overall quality assessment for lecturers’ assessment in Malaysia and in their means: from a locally developed self-assessment test in Germany, to reference of paneuropean Common European Framework of Reference (CEFR) levels in Indonesia, Denmark and Germany.

<table>
<thead>
<tr>
<th>Country</th>
<th>Assessment tool</th>
<th>Criteria</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malaysia</td>
<td>Staff performance appraisal</td>
<td>Publication/Writing Research Conference</td>
<td>Salmuni et al. (n.d.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teaching Consultancy Services Personality</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>SoTL ‘movement’</td>
<td>Structured programme leading to publications</td>
<td>Plews and Amos (2020)</td>
</tr>
<tr>
<td>Denmark</td>
<td>TOEPAS</td>
<td>TOEPAS linked the to the CEFR criteria</td>
<td>Kling and Stæhr (2012)</td>
</tr>
<tr>
<td>Indonesia</td>
<td>Postgraduate program in English instruction-based universities</td>
<td>English for Academic Purposes test score must reach international level to at least CEFR level B2</td>
<td>Marsaulina (n.d.)</td>
</tr>
</tbody>
</table>
As we can see from Table 1, three out of five cases (Denmark, Germany and Indonesia) refer to CEFR levels B1 and B2, focusing on speaking skills, while the other two (Australia and Malaysia) stress the importance of writing and its link with the lecturers’ ability to produce publications.

**ASSESSMENT CRITERIA FOR MEDIATION**

Although the *Companion Volume of the Common European Framework* (Council of Europe, 2018) has the same system of language assessment levels as CEFR (Council of Europe, 2001), we can see a shift from the traditional division between writing and speaking skills to the integration between reception, production and interaction strategies, providing many new scales for mediation strategies. These are defined as follows: ‘In mediation, the user/learner acts as a social agent who creates bridges and helps to construct or convey meaning, sometimes within the same language, sometimes from one language to another’ (Council of Europe, 2018: 103). The creation of new meaning is mentioned in the definition above, but the construction of new knowledge is added when discussing ‘mediating functions: that of organising collective work and the relationships between participants; that of facilitating access to – and the construction of knowledge’ (ibid.: 43).

According to Hyland (2015: 3), organising collective work and constructing knowledge directly relate to the essence of the work of academic staff: ‘Academic publishing is central to the construction of knowledge and the measurement of the academic’s professional competence’. Thus, the mediation construct allows us to connect the Malaysian and the Australian view of academic performance first and foremost as the production of publications with the Danish, German and Indonesian academic staff assessment systems, focusing on the lecturers’ spoken performance (see Table 1 above).

The choice of mediation descriptors for assessment of academic staff can also be motivated by the description of a successful mediator at level C1 (Council of Europe, 2018: 105), as someone who (1) helps to maintain positive interaction by interpreting different perspectives; (2) manages ambiguity, anticipates misunderstandings; (3) builds on different contributions to a discussion; (4) stimulates reasoning with a series of questions; and (5) conveys clearly and fluently in well-structured language the significant ideas in long, complex texts (Council of Europe, 2018: 105).

All these activities can be seen as desirable in a lecturer, as we are navigating the misunderstandings between the traditions of our scientific field and the student interpretation of its demands, organising discussions, trying to stimulate the thinking of the students and trying to convey the latest research as clearly as we can.
RESEARCH METHOD

Construct validation method has been recently analysed by Im, Shin and Cheng (2019), where they compare the approaches proposed by Messick (1989), Mislevy, Steinberg and Almond (2003), Chapelle et al. (2008) and Kane (2013). In the present study, we follow Chapelle et al.'s (2008) framework, as we focus on mediation strategy use, which we have described and operationalised in our test, observed the score obtained by the test takers, compared it to the target score and used the obtained results to assess the test takers’ performance.

Figure 1  **Construct validation methods** (Im et al., 2019)

Compared to Bachman and Palmer’s *Assessment Use Argument* (2010) and their focus on interpretation, decisions and consequences, Chapelle et al. (2008) are concerned with the theoretical and empirical construct validation, which is more appropriate in our case, as we can control the theoretical construct, but cannot control the decisions taken by the administration of the university on the basis of the administered test. Chapelle et al.’s (2008) approach also allows us to integrate the quantitative and qualitative validation methods using a mixed method approach.

RESEARCH CONTEXT

The University of Latvia started the academic staff competence enrichment project in 2018. The aim of the project is to promote the professional development of the UL staff, to attract young and talented international as well as local researchers. One of the objectives of the project is to develop the academic and
professional English language skills of the academic staff in the humanities, social, medical and life sciences for their use in the academic setting.

Assessment is an integral part of language teaching in the project. The course starts and ends with English language proficiency assessment using standardised tests aimed at levels B1-C1. The test administered before the study course is based on an international placement test using discrete multiple-choice items to assess lexical and syntactical structures. The test was administered online in Moodle, the test takers were not observed; thus, the tests can be considered as a form of self-assessment. The test results were used to divide the academic staff into B1, B2 and C1 groups. The total volume of the course comprised 210 academic hours providing the staff with the necessary professional English language support for the acquisition of lexical and grammatical knowledge, spoken and written proficiency as well as reading and listening skills through teaching a blended learning multimodal course adjusted to the specific needs of each specific group. The staff were assessed again at the end of the course.

The end of the course test was developed by the English language teachers of all the groups, then the tests were pretested in other academic staff groups and analysed using Iteman (ITEMAN Classical Test Theory Analysis, n.d.). The items with positive discrimination, and tasks whose alpha coefficient was not lower than 0.6-0.7 were edited and combined in a single test, ranging the difficulty levels from B1 level to C1 level, from separate reading and listening skills to integrated reading/writing skills tasks. The test specification can be seen in Table 2.

The speaking test was administered by two people, both taking notes and assessing the performance after each case and assigning the points for presentation skills, interaction skills, data interpretation skills, linguistic range and phonological control. The multiple choice and the matching tasks were clerically marked, but all the other tests were marked using marking scales, which were updated while marking in order to incorporate the details of performance and improve the reliability and validity of the assessment. All the results were data processed, and their reliability and internal consistence were checked by calculation of correlation.

**TEST RESULTS**

After the delay caused by Covid 19 epidemics, the test of 2020 was taken by 68 academic staff members who had followed the course for a year. The test construct is based on equal distribution of 20 points for each language skill (reading, writing, listening and speaking). It starts with a listening task of a recorded fragment of a two-minute presentation, where the test takers fill in the gaps in the transcript of the recording with such words as feedback, components, focus, interviewing, tracking. The aim of the task is to assess the test takers’ detailed understanding of the audio recording and their ability to identify important academic vocabulary in the recording. This was aimed at level B1 test takers,
and it was also the easiest task with 78 per cent average, compared to the reading multiple choice task, where the facility value was 51 per cent. The difference of 27 per cent between the easiest and the most difficult items supports our claim that the test is appropriate for a range of performance levels (B1-C1).

### Table 2 Academic staff test specification, task weighting and mean

<table>
<thead>
<tr>
<th>Level</th>
<th>Skill</th>
<th>Task description</th>
<th>Maximum Points</th>
<th>Mean %</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>Listening for detail</td>
<td>Gap fill</td>
<td>6 points</td>
<td>78</td>
</tr>
<tr>
<td>B1-B2</td>
<td>Listening and writing (Spoken text mediation)</td>
<td>Summarize the lecture and compare the situation to your own</td>
<td>20 points</td>
<td>53</td>
</tr>
<tr>
<td>B2</td>
<td>Reading</td>
<td>Multiple choice task</td>
<td>4 points</td>
<td>51</td>
</tr>
<tr>
<td>C1</td>
<td>Reading and writing (Written text mediation)</td>
<td>Text summary in a sentence</td>
<td>10 points</td>
<td>58</td>
</tr>
<tr>
<td>C1</td>
<td>Reading</td>
<td>Arrange five sentences in a paragraph</td>
<td>10 points</td>
<td>59</td>
</tr>
<tr>
<td>B1-C1</td>
<td>Speaking (Spoken data interpretation)</td>
<td>Present yourself, answer questions about your research, describe a graph</td>
<td>20 points</td>
<td>73</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>70</td>
<td>66</td>
</tr>
</tbody>
</table>

The second task is also a recording of a three-minute fragment of a lecture on the change of the prestige of the university degree. The test takers can take notes while listening, and the task is to mediate between the spoken and written medium, summarizing the lecture and comparing the lecturer’s views with their own. This task was assessed using five different criteria from the CEFR CV. Descriptors for Processing text in speech (Council of Europe, 2018: 111) and grammatical accuracy (Council of Europe, 2001: 114).

The third task is a multiple-choice reading task, consisting of four paragraphs, each followed by a multiple-choice question with three distractors. This is followed by a text summary task, which will be discussed in detail later, a rearrangement task and a speaking test, where the test takers needed to describe their research profile and answer questions, describe a graph, which was assessed using the descriptors from CEFR CV (Council of Europe, 2018), namely, presentation skills, interaction skills, data interpretation skills, linguistic range and phonological control.
The minimum performance of the whole test was Min: 23 per cent, Max: 97 per cent, St. dev: 10.19, Mode: 64 per cent, Mean: 66 per cent. Figure 2 shows the distribution curve in the absolute numbers, which has a slightly negative skew.

**Figure 2** Histogram of the test-taker performance in the post teaching test

The cut off points were primarily decided using the mean and the standard deviation (St. dev. 10.19), then the test scripts near the cut off points were examined carefully, and the borders of the levels were shifted to ensure the reliability of assessment, see Figure 2. The cut off points were established on the basis of standard deviation to convert the points into levels (see Figure 3).

**Figure 3** Histogram of levels B1, B2, C1 and their cut off scores

The test results were validated using qualitative and quantitative validation methods. **Qualitative validation:** once the test takers were assigned their levels, those were discussed with the teachers of each group to ensure concurrent validation. The test results and the scripts were also discussed with the test takers themselves in case they felt that the result did not agree with their self-assessment (response validation), or they wanted to discuss the results for further improvement.
Quantitative validation: all the test results were checked using correlation coefficients, which were all positive, although the tasks that assessed reading and listening skills separately had consistently lower correlation levels. The correlations of the test results suggest that the tasks that were aimed at text mediation had the highest internal consistency (see Table 3 below): the correlation between the listening gap filling task and the spoken mediation task (summary of a lecture and own interpretation) is significant but not very high (0.523), although both tasks were based on a spoken text input. The correlation between the written text mediation (summary of a written text) and spoken text mediation (summary of a spoken text and its interpretation) is higher (0.643). The spoken text mediation has also a significant correlation with spoken data interpretation (0.694). The highest correlation index, however, can be observed when correlating the total written test performance with the total results (0.946), suggesting the importance of the combination of reception, production, interaction and mediation strategies for more reliable and valid test results. The low correlations (Cronbach Alpha test) for the sentence arrangement task and reading multiple choice tasks suggest reliability issues for the tasks, which could have been caused by the high facility value for the listening task (see Table 3, but not for the reading sentence arrangement task, which asks for further analysis, which is outside the scope of this research.

Table 3 Correlation indices between the separate tasks

<table>
<thead>
<tr>
<th></th>
<th>Listening gap fill</th>
<th>Spoken text mediation</th>
<th>Reading multiple choice</th>
<th>Written text mediation</th>
<th>Reading sentence arrangement</th>
<th>Total written</th>
<th>Spoken data interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening gap fill</td>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spoken text mediation</td>
<td>0,523*</td>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading multiple choice</td>
<td>0,116</td>
<td>0,379*</td>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written text mediation</td>
<td>0,573*</td>
<td>0,643*</td>
<td>0,383*</td>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading sentence arrangement</td>
<td>0,085</td>
<td>0,177</td>
<td>0,197</td>
<td>0,283</td>
<td>1,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total written</td>
<td>0,662*</td>
<td>0,895*</td>
<td>0,501*</td>
<td>0,824*</td>
<td>0,456*</td>
<td>1,000</td>
<td></td>
</tr>
<tr>
<td>Spoken data interpretation</td>
<td>0,421*</td>
<td>0,694*</td>
<td>0,453*</td>
<td>0,547*</td>
<td>0,180</td>
<td>0,687*</td>
<td>1,000</td>
</tr>
<tr>
<td>Total</td>
<td>0,633*</td>
<td>0,873*</td>
<td>0,503*</td>
<td>0,784*</td>
<td>0,369*</td>
<td>0,946*</td>
<td>0,845*</td>
</tr>
</tbody>
</table>

*Statistically significant correlations
MARKING CRITERIA FOR MEDIATION OF TEXTS

The marking criteria for the integrated listening to a lecture and writing a summary task were developed based on the Companion volume of CEFR (Council of Europe, 2018: 111) using receptive, productive, interactional and mediation strategy scales mediating a text: processing a text in speech. These were used to assess the test takers’ ability to relay the spoken text in writing, using Jeffrey Selingo’s speech on the Value of a College Degree (2016) as input.

Table 4 presents a fragment of the marking scale CEFR (ibid.) with the first sentence of the sample scripts for each level. As we can see, the first script (4 points) starts their summary by stating not only the topic, but also the speaker’s main message, which further serves as the main hypothesis to be argued using the details of the speech. The second script (3 points) states the topic of the speech precisely enough, which is the value of education, but does not reveal the message of the speech. The third script (2 points) describes the topic partly, thus including incomplete information that interests them more, namely about the attitude of students instead of society, while the fourth script (1 point) claims that the speech is about the education system and many countries in general, thus failing to connect to the original lecture and its topic of the change of the value of college degree.

Table 4  Mediation strategy assessment scale with sample script fragments with the original spelling

<table>
<thead>
<tr>
<th>Points</th>
<th>Processing text in speech</th>
<th>Sample script (The first sentence of the summary)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Clear, complete summary of essential information.</td>
<td>According to the presenter the value of the university degree has decreased. While in the past a degree was perceived as a guarantee of prestigious and well-paid employment, now it may just be enough to get any job.</td>
</tr>
<tr>
<td>3</td>
<td>Summarises the important points; one essential point may be missing.</td>
<td>Jeff Selingo gives us his point of view how valuable is a college degree today and in the future. In the beginning of the presentation he mentioned that degree was a signal that you were ready to work.</td>
</tr>
<tr>
<td>2</td>
<td>Includes incomplete information. Some important points are missing.</td>
<td>The topic was about new graduated and their problems to find appropriate job. Jeff Selingo thinks that a lot of students have not correct attitude to studies.</td>
</tr>
<tr>
<td>1</td>
<td>Information does not connect to the original lecture.</td>
<td>This recording illustrates education system. In many countries was observed that the university degree was not for a good job.</td>
</tr>
</tbody>
</table>
DISCUSSION OF THE TEST RESULTS

Mediation strategy inclusion in the Academic staff English language test and its validation process can be supported by theoretical and empirical arguments:

First, mediation descriptors from the CEFR companion volume (Council of Europe, 2018) agree with the everyday tasks of the academic staff, as we need to mediate between the student knowledge and the theoretical knowledge of our research area which is available in spoken and written texts as well as visual information graphs in English language medium instruction. The mediation descriptors can also be used in assessing academic staff’s written and spoken performance for both objective and subjective tasks.

Second, Chappelle et al.’s (2008) construct validation framework allows us to examine the construct of a language test from the very beginning of test conception to the test-taker performance analysis and its interpretation, thus enabling us to develop a test, evaluate the performance, extrapolate the findings and generalise back to the construct measured by the test.

Finally, both quantitative and qualitative analyses of the English language performance of the Academic staff suggest that the mediation strategies enrich the construct with variables not tested by the traditional reading, listening, writing and speaking tests as suggested by the correlation indices between the different tasks in the academic English performance test. The correlation indices suggest inner construct existence that needs further research using, for example, structural modelling methods.

CONCLUSIONS

Internationalization of higher education has implications for the use of English in knowledge generation and dissemination in EMI programmes. Successful implementation of EMI programmes requires addressing the needs of all stakeholders, including qualified academic staff, who are required to know the English language at least at level B2, preferably at C1 level. Thus, assessment of academic staff is carried out across the world. The systems and the criteria differ, but there are also common variables: spoken and written production is the preferred media for assessment, while mediation strategies are not so often used. Nevertheless, the mediation strategies construct is similar to the role of the academic staff as new knowledge constructors creating new meaning to texts in different languages and different media. Mediation strategy construct operationalization in the new descriptors of the CEFR allow us to assess the ability of the academic staff to mediate between the spoken and written media, and the level descriptors allow us to discriminate between the different levels of ability to manipulate the meaning of the spoken text in a written medium.
REFERENCES


**Vita Kalnbērziņa** (Dr. Philol.) is working as an associate professor at the University of Latvia. Her research is mainly focused on language acquisition and assessment as well as intercultural communication and bilingualism. Email: Vita.Kalnberzina@lu.lv

**Indra Karapetjana** (Dr. Philol.) is working as a full professor in applied linguistics at the University of Latvia. Her main research interests involve political and academic discourse analysis, critical discourse analysis, ESP and CLIL. Email: Indra.Karapetjana@lu.lv