Language Assessment of EMI Content Teachers: What Norms

Dimova, Slobodanka

Language Assessment of EMI Content Teachers: What Norms

Abstract

A number of universities have implemented policies for internal assessment of lecturers' English

proficiency to ensure the quality of teaching in EMI programs. Given the complexity of the

multilingual teaching and learning contexts in which these assessments are administered and

used, decisions about what English norms lecturers' performances are assessed against have been

a struggle. The selection of appropriate norm can warrant the quality of the assessment

instrument and the validity of the procedure, as well as the avoidance of bias. Therefore, this

chapter discusses the selection of English norm(s) in relation to the validity of a performance-

based EAP certification test.

Author bio

Slobodanka Dimova is an Associate Professor in Language Testing and the coordinator

of the Test of Oral English Proficiency for Academic Staff (TOEPAS) at the University of

Copenhagen, Denmark. She serves as the Book Review Editor for the journal *Language Testing*,

a member of the Executive Committee of the European Association of Language Testing and

Assessment (EALTA). Her research interests include rater training and behavior for speaking

tests, development and validation of performance-based tests scales, use of technology in

language testing and assessment, as well as policies and practices concerning the implementation

of English-medium instruction (EMI) programs at non-Anglophone universities.

Language Assessment of EMI Content Teachers: What Norms

1. Introduction

Internationalization has become one of the main strategic goals of universities as part of their endeavours to prepare students for careers in a globalized world. Although curriculum and program internationalization may be one way to implement this strategy, many universities internationalize through recruitment of international students and lecturers. Therefore, apart from university participation in exchange programs (e.g., Erasmus) and joint degrees, which increase student and lecturer mobility, full-degree international student recruitment and international lecturer employment are also widespread (Mauranen, 2010). Regardless of which internationalization strategy universities use, English seems to play a central role in the process, as many non-Anglophone universities have been implementing English medium instruction (EMI) courses and programs (Wächter & Maiworm, 2014). It has been widely accepted that EMI allows for easier recruitment of international students and lecturers who are more likely to know English than the local language, as well as access to research, textbooks, and other teaching materials that are published outside the local context. Given the linguistic diversity of teachers and students in the university contexts, EMI refers to a multilingual university environment in which English is used as a lingua franca in academia (ELF in academia).

The implementation of EMI courses and programs at non-Anglophone universities has also raised concerns about the quality of teaching and learning, given that most stakeholders involved in the process are L2 speakers of English (Klaassen & Bos, 2010; Kling & Hjulmand, 2008; Kling & Stæhr, 2011). For that reason, alongside internationalization policies, a number of universities across Europe (e.g., University of Freiburg, Aarhus University, Maastricht University, University of Nantes, Complutense University) have also implemented policies for internal assessment of lecturers' English proficiency with the intention to ensure the quality of

teaching. Given the linguistic complexity of the local teaching and learning situations in which these assessments are administered and used, the selection of language norms lecturers' performances are assessed against may not be straightforward and simple. Therefore, many assessment methods rely on native-speaker norms, such as standard American or British English, despite recognitions of their inadequate applications in an ELF in academia context (Mauranen, 2012; Murata & Iino, 2017; Smit, 2010). Notwithstanding the technical restrictions to capture the linguistic complexities of ELF in academia in a language assessment situation, this chapter points the found inadequacies of native-speaker norms for oral English proficiency tests in EMI contexts and provides recommendations for a more valid language assessment.

More specifically, the chapter presents the findings from a study investigating the role of norms in a locally-developed assessment method used to certify EMI lecturers for oral English proficiency at the University of Copenhagen (UCPH), the Test of Oral English Proficiency (TOEPAS). The study was guided by research questions related to: a) frequency of references to language norms in the TOEPAS written feedback reports, and b) EMI lecturers' perceptions of their own English language proficiency and their communicative classroom practices. The findings from the study had implications for the development of scalar descriptors and feedback formats of the second version of TOEPAS, which are discussed in the final sections of the chapter.

2. English Medium Instruction: Definitions of contexts and language uses

Considering that internationalization in higher education results in multilingual classroom settings in which English is the medium of communication inside and outside the classroom, two sociolinguistic phenomena relevant for such settings provide the foundation for the discussions

in this section, namely ELF in academia and translanguaging. Before these two phenomena are examined regarding possible applications in language assessments that measure the ability to communicate in the EMI classroom, the definition and characteristics of EMI are provided.

EMI can be defined as a situation where, "non-language courses in for instance medicine, physics, or political science are taught in English, to students for whom it is a foreign language. As often as not, it is also taught by a lecturer who does not have English as a first language (L1)" (Hellekjaer, 2010, p. 11). Reviewing different EMI definitions, Pecorari and Malmström (2018) concluded that EMI is commonly described applying four main characteristics:

1. English is the language used for instructional purposes.

2. English is not itself the subject being taught.

3. Language development is not a primary intended outcome.

4. For most participants in the setting, English is a second language (L2). (p.499)

The fourth characteristic implies that in EMI, English is used as ELF in academia, where underlying multilingualism among the participants exists. Research based on spoken academic ELF speaking corpora (Mauranen, Hynninen, & Ranta, 2010) suggests that alongside the structural variations of the language at morphological and phrasal levels (Mauranen, 2007; 2012; Ranta, 2006), the central communicative practices that characterize the ELF in academia discourse in the multilingual contexts are negotiations of meaning and practices of mediation (Hynninen, 2011; Mauranen, 2010). In other words, participants in ELF in academia tend to negotiate meaning to a higher degree than in monolingual situations in order to achieve intelligibility. This negotiation may involve topic negotiation, meta-discourse, and rephrasing in

order to achieve common ground with interlocutors who may lack the same cultural and/or educational background.

When it comes to teaching in ELF in academia contexts, Suviniitty (2012) found that the interactional features, rather than perceived English skills, improved the accessibility of lectures for students in the international classrooms. Björkman (2010, p. 88) lists a number of pragmatic strategies lecturers need to apply claiming, "effectiveness in ELF settings seems most strongly associated with one's pragmatic ability and less with one's level of proficiency". Some of the pragmatic strategies Björkman (2011) found particularly useful in teaching were to: 1) name the focal processes, 2) mark major points explicitly, 3) establish cohesion and coherence through signposting and repetition, and 4) organize topics.

The underlying multilingualism among the participants, on the other hand, leads to opportunities for using multiple codes in the classroom. As Pecorari and Malmström (2018) note, research has revealed different modes of multiple-code application in EMI classrooms, ranging from insertion of L1 terminology in predominantly English communication to situations where on-record communication (explaining terminology, concepts, phenomena) is in English and off-record communication is in L1, if the local language is shared. These shifts between different linguistic repertoires could be referred to as "translanguaging" (Garcia & Li, 2014), where participants flexibly utilize their own linguistic repertoires. The linguistic resources students possess can serve as "affordances for learning," where the language systems in the speaker's linguistic repertoire are not delimited but negotiated in relation to particular situations in which communication occurs (Canagarajah, 2016, p. 267). Referring to EMI, Coleman (in Coleman, Hultgren, Li, Tsui, & Shaw, 2018) claims that translanguaging in an ELF setting allows fluid movement between different languages, but also between varieties of the same language.

Translanguaging supports the process of successful learning by drawing on and mediating multiple linguistic resources (Garcia & Li, 2014; Li, 2011; Joe & Lee, 2013; Yeh, 2012). In EMI, "translanguaging allows exploration of the nonnative uses of English for communication purposes and also acknowledges the existence of the English varieties residing in multilingual settings" (Coleman et al., 2018, p. 715). The dilemma, though, is how to elicit the translanguaging space in language assessment situation when translanguaging tends to occur spontaneously, and the "repertoires can only be understood by attending to their functions, i.e., to their actual and contextual deployment, not to any abstract or a priori assessment of what they mean or of what they are worth" (Blommaert, 2015, p. 15). To include translanguaging as part of the EMI communication construct requires development of assessment tasks that elicit observable translanguaging practices. However, given the assumptions that the language resources used in a particular context differ among language users, and that translanguaging is spontaneous rather than predictable, then designing adequate assessment tasks that elicit translanguaging remains a daunting effort.

In light of these discussions about universities as multilingual ELF contexts, the existing language assessment methods used for measuring students' or lecturers' language proficiencies have been criticized for failing to represent the complex linguistic realities of the context (Chopin, 2014; Jenkins & Leung, 2017; Newbold, 2014). Language testing has been criticized for overuse of native-English-speaker norm criteria and reliance on language standards that dismiss the communication patterns of English spoken in many contact situations (Davies, Hamp-Lyons, & Kemp, 2003). Considering the diversity and multiplicity of contexts in which language tests are used, the local validity of assessment tasks based on a standard norm becomes problematic if the standard norm deviates from the norm(s) used in the local setting (Lowenberg,

1993). Critics have pointed out that language assessments lack methods that measure the ability to translanguage, so assessment traditions diverge from research that provides evidence that translanuageing is "the most significant indicator of linguistic multicompetence, and a central contributing factor to the cognitive reserve that ultimately gives the bilingual and multilingual the cognitive and social advantages" (Coleman et al., 2018, p. 709-10). Instead of translanguaging, "an idealized native speaker norm" continues to permeate English language assessment for EMI (see Li in Coleman et al., 2018, p. 709). Critics have also firmly maintained that as major sites for ELF communication, it is essential for universities to develop "user-centred and norm-defocused" assessment that allows for more variation and reflects the pragmatics of ELF interaction as well as the patterns of effective communication and negotiation of meaning (Chopin, 2014; Newbold, 2014). Despite the acknowledgement of the research and the development of multilingual paradigms in sociolinguistics and pragmatics, including multiple norms or languages in the design of relevant tasks remains a challenging task due to the current constraints guiding the development of valid tests that represent the domains of target language use (Dimova, 2017a; Elder & Harding, 2008, Harding & McNamara, 2017). As Shohamy (2017) suggests, ample research needs to be performed to investigate how the different contexts and the underlying speech production processes in ELF, translanguaging, and other bi- or multi-lingual modalities can be adequately represented in language tests (p.591).

Assessment of lecturers' English proficiency, in particular, is deemed inappropriate when the assessment criteria are based on the linguistic aspects of EMI lecturers' speaking performance in a mock lecture (Coleman et al., 2018; Macaro, 2018). Suggestions have been made that these assessments should focus on pedagogy instead of language. However, while on the one hand such assessments may contextualize the language use, focus on pedagogy may be

discriminatory because lecturers' pedagogy would be scrutinized only if they teach in English and not in other languages (Dimova & Kling, 2018). In other words, current discussions regarding lecturer assessment tend to conflate lecturers' classroom behaviour (behavioural pedagogy) and the language tools they use to communicate content material to students (linguistic pedagogy). Assessment of lecturers' language resources and strategies for effective communication in a multilingual classroom should remain in the focus. However, the type of classroom activities and materials lecturers use are often grounded in the teaching traditions of the field or the local university and may need to be addressed separately.

3. Methods for assessing lecturers' language

When developing language assessment methods, it is important to investigate the target language use domain (TLU), which is defined as, "a set of specific language tasks that the test taker is likely to encounter outside the test itself, and to which we want our inferences about language ability to generalize" (Bachman & Palmer, 1996, p. 44). Namely, it is necessary to identify the specific communicative tasks in which test takers engage in real life and analyse the successful language uses pertinent to these tasks so that the same task types and the same language functions can be elicited with the test (Bachman & Palmer, 2010). With reference to development of methods for assessing lecturers' language uses in EMI, identifying the TLU means understanding the tasks lecturers perform in the teaching domain, or other domains deemed necessary (e.g., mentoring, meetings, online communication). In addition, the characteristics of language use in the EMI domain also need to be described in order to serve as a yardstick against which test performances will be measured.

Although the teaching tasks may be easily identifiable, the intricacies of language uses in EMI may not be described and coded with the same degree of ease because our language conceptualizations lie entrenched in well-defined, monolithic, compartmentalized systems (Shohamy, 2011). Söderlundh (2013), who defines norms as explicit and implicit "expectations concerning social and linguistic behaviour," has found that in EMI classroom settings in Sweden, various general and group-specific norms exist depending on students' and teachers' linguistic resources. While English is generally accepted as the shared language in formal classroom communication, the predominant local language and other languages are used among different student groups in the classroom. These norms are emergent and transient because they shift depending on the users and the context (Mortensen, 2014).

Therefore, the EMI contexts and their communicative patterns may vary based on different underlying factors (e.g., students' and lecturers' linguistic repertoires, their educational/cultural backgrounds, difficulty of subject content). To what extent the test method can elicit relevant language on the basis of which lecturers' communicational effectiveness can be generalized stays questionable. For example, how to operationalize the structural variations observed in ELF in academia or the realizations of translanguaging represents a challenge as they are unevenly distributed and deeply embedded in the particular occurrence of the contextual space. In other words, it is difficult to identify specific patterns of how to balance linguistic variation in lectures to accommodate the huge linguistic heterogeneity of the EMI student population.

Solano-Flores (see Menken & Shohamy, 2015) proposed some strategies to promote translanguaging-based methods to account for linguistic variation for assessment in linguistically diverse contexts, which include test localization, psychometric examination of linguistic

heterogeneity, and analysis of the relationship between language and the other constructs (e.g., pedagogy) that are measured by tests. Applying these strategies in the EMI context, however, is not an easy feat because the linguistic heterogeneity of the teacher and student population tends to be quite fluid, compared to that of local bilingual communities. In other words, the code-switching patterns in bilingual communities, i.e. communities who switch between two different codes because they share the same two languages, may be more stable and predictable than the transient communicative norms developed in multilingual groups.

Nevertheless, attempts have been made to develop language assessment methods that include aspects of translanguaging in the university context (Baker, 2017). To reflect the realities of translanguaging practices at a large Canadian university, listening tasks that include shifts between English and French were designed as part of an assessment tool for university lecturers. The assessment was developed as a response to the university's bilingual language policy that requires lecturers to demonstrate "passive" bilingualism for the purpose of promotion and tenure (Baker, 2017).

These attempts to include translanguaging in formal assessment contexts allow the field of language testing and assessment to keep abreast with the current discussions in sociolinguistics. However, although designing receptive skill items (reading and listening) based on authentic texts from the translanguaged use domain for which generalizations will be made seems viable, designing productive skill items (writing and speaking) that assess lecturers' translanguaging skills may be a less obvious endeavour. The challenges associated with assessing lecturers' translanguaging skills for an EMI program stem from the lack of rigorous operationalization of the translanguaging construct. Clear understanding is missing with regard to what represents the translanguaging space, how the linguistic repertoire is activated, and, more

importantly, how appropriate translanguaging can be elicited in a testing situation, and what valid generalizations can we make based on the test performance.

Given the technical difficulties in the design of reliable and valid assessment tools that embrace the linguistic complexities of the EMI settings, lecturer EMI certification tends to be based on uses of general academic English tests or locally-developed assessment methods that are implicitly based on standard Englishes (Ball & Lindsey, 2013; Klaassen & Bos, 2010; University of Freiburg, 2018). For example, although institutions recognize the multilingual, multicultural university contexts, the rating criteria may focus on accurate pronunciation and lexical use with implied native-English-speaker norm references (University of Freiburg, 2018).

In a similar manner, when first developed, the Test of Oral English Proficiency for Academic Staff (TOEPAS), which is used to certify EMI lecturers for oral English proficiency (Kling & Stæhr, 2011; 2012), included the *educated native speaker* reference at the highest scalar levels, which is an explicit reference to a standard variety. Taking into consideration the ELF in academia contexts of the EMI programs at the University of Copenhagen (UCPH), this descriptor was intended as a point of reference, rather than a norm against which performances were assessed. Nonetheless, to warrant the validity of the assessment instrument, it was important to examine how native-speaker references, be they explicit or implicit, affect the outcomes of the assessment process. The results of such examination indicated in which direction the revisions of the assessment instrument would be developed to achieve a better alignment with the intended assessment outcomes, as a monolithic norm may be deemed inacceptable for the EMI context.

4. Methods

To examine the effects of the native-speaker norms in the TOEPAS, the oral English proficiency certification for EMI lecturers at UCPH, the current study was guided by the following questions:

1. How are native-English-speaker references used in the TOEPAS 1.0 written feedback reports written by the raters?

2. How do EMI lecturers perceive their own language proficiency and the use of English in the EMI classroom after receiving the feedback?

The first question deals with the explicit uses of the native-speaker norms in the TOEPAS written feedback reports, and the second question elicits the implicit influences of these norms on lecturers' perceptions of English language proficiency and EMI classroom communication. To answer the research questions, the content of the TOEPAS written feedback reports (n=400) was analysed, and interviews with lecturers who had taken the TOEPAS were conducted (n=24). In order to contextualize the TOEPAS feedback reports and lecturers' opinions, the following section provides background information about UCPH and TOEPAS.

4.1. The Test of Oral English Proficiency (TOEPAS) at the University of Copenhagen

In 2008, the management at the UCPH decided to implement an assessment procedure to certify the oral English skills of the lecturers who would teach in select graduate EMI programs at the university. The desired purpose of the certification would be twofold: 1) assess whether the lecturers had the necessary oral English language skills to cope with the communicative demands of teaching in EMI programs and courses, and 2) identify lecturers' language needs and inform them about suitable language support or training.

In order to design such a testing method, the target language use (TLU) domain was identified by analysing the specific language tasks the EMI lecturers encounter in the real-life context (Kling & Stæhr, 2012). This step involved course observations, interviews with lecturers, deans, and heads of study boards. Once the TLU domain was analysed, the TOEPAS tasks were developed to reflect the key language characteristics found in the TLU domain so that the test-takers' performances could be interpreted as evidence of their ability to perform similar language tasks beyond the testing situation. This would ensure the authenticity and the (construct) validity of TOEPAS (Douglas, 2000, p. 47).

The TOEPAS format presents a simulated teaching performance in a controlled setting. Each TOEPAS test administration lasts roughly two hours and involves the assessment of three lecturers from the same program or area of expertise. Test-takers take turns giving a prepared mini-lecture and participating in a role-play as 'students' in order to simulate a graduate classroom setting. As 'students', the participants are asked to both interrupt the lecturer (test-taker) during the lecture and to ask spontaneous questions after the lecture. The shared expertise of the lecturers allows them to engage in meaningful interaction about the specialized topic. The test-takers' lectures are based on a subject or topic that they typically teach.

All TOEPAS administrations are digitally video recorded for use in both assessment and feedback. With the TOEPAS 1.0 version, two trained raters rated the test takers' performance on a 5-point holistic scale taking into account their observation of the live performances (see TOEPAS scale grid in Dimova, 2017b, p.57). The TOEPAS test takers ultimately receive a certificate with their holistic result, as well as a detailed feedback report, accompanied by a digital link to their recorded performance. The feedback report of the TOEPAS 1.0 version provided description of test-takers' speaking performance with an assessment grid that focused

on five main areas: fluency, pronunciation, grammar, vocabulary, and interaction. Quotes from the actual language observed in the test-takers' performance were used to exemplify the feedback. For example, if the feedback stated that the test-taker made mistakes with congruence in grammar, then actual examples of these mistakes would be provided. Research suggested that despite the technical terminology, the test-takers (lecturers) valued these examples because they were concrete, and they could easily remember them (Dimova, 2017b).

4.2. Written formative feedback analysis

To answer the question about the uses of native-speaker references in the written formative feedback reports, all written reports from the TOEPAS database were included in content analysis (n=400). Content analysis is a process of categorizing textual data into clusters to identify consistent patterns and relationships between keywords or themes that represent certain variables (Julien, 2008: 120-122). To analyse the textual data in the written reports, a keyword search in NVivo 10 was performed, and the frequency of the words "native" and "non-native" and the word clusters around these keywords were analysed. The keyword/key text search allowed for comparison of the recurrence of native-speaker references across the different TOEPAS levels represented in the written reports, as well as the most common textual environment in which they occurred.

4.3. Semi-structured interviews

In-depth semi-structured interviews (Seidman, 2013) were conducted with randomly selected academic staff members who had taken the TOEPAS to collect data regarding lecturers' opinions about language uses and EMI classroom practices. Semi-structured interviews seemed most appropriate to answer this question because they would allow to understand EMI lecturers'

cognition (see Borg, 2009) with respect to the meaning of the TOEPAS results and their reported practices by using a specific set of questions and through exploration of new ideas and themes which may come up during the interview (Mason, 2004).

The interviews were designed and conducted in three stages: (1) interview question design, (2) procedures and questions piloting, and (3) interview administration. The interviews, which lasted around 30 minutes, were audio recorded and subsequently transcribed, coded, and analysed with NVivo 10. The coding procedure began with data segmentation based on turns as individual segments because participants rarely shifted topic without being prompted by a question.

The final semi-structured interview consisted of 23 basic questions related to four different themes: information about teaching experiences, video from test performance, written report with formative feedback, and perceived test impact. Additional questions were raised during interviews to further explain certain statements or to obtain more details regarding specific cases. Instead of direct questions about native vs. non-native speaker or standard vs. deviant norms, questions focused on lecturers' perceptions of classroom communication and their own language proficiency. The reason for the indirect approach was to find out whether references to norms would transpire in lecturers' responses when they discuss language behaviours in the classroom. Therefore, questions sought information related to participant's experience with teaching in EMI programs, international students, communication practices in and outside the classroom. The video and written report questions related mostly to whether and how participants used them to improve their language skills for the classroom, while the last set of questions inquired about the impact of TOEPAS on the individual participants, i.e. what the TOEPAS results meant for them.

4.4. Participants

Participants were 24 teaching staff members (assistant, associate, and full professors, as well as post-docs and PhD students) who took part in the certification process across different faculties and departments at the University of Copenhagen (6% of the entries in TOEPAS 1.0 database). They were randomly selected from each scale level in the TOEPAS 1.0 database. The database which was designed in MS Access, allowed for random selection queries. If the selected participant could not be reached after two e-mails and five phone calls, because some lecturers had left the university or were on leave, new stratified database queries were performed to ensure random selection of participants. During the pilot, it became evident that participant recruitment was easier by phone than e-mail, so all participants were telephoned for invitation and scheduling of the interview time. Table 1 shows the background characteristics of the participants.

Years at UCPH	range=1-36 years, median=11
Gender	F=5; M=19
L1	Danish=19; Swedish=2; Portuguese=1; Dutch=1; German=1;
Department	Computer Science Department of Forest and Landscape Institute of Food and Resource Economics Department of Agriculture and Ecology Department of Food Science Department of Large Animal Science Department of Human Nutrition Law
Position	Full Professor (n=6); Associate Professor (n=13); Assistant Professor (n=2); Post-doc (n=1); PhD student (n=2)
Score	5 (n=4); 4 (n=10); 3 (n=9); 2 (n=1)

Table 1: Participant background information

5. Results

     The content analysis of the TOEPAS feedback reports and the semi-structured interviews yielded complementary data, which assisted the overall interpretation of the findings. The following are the findings from each analysis and the overall implications.

5.1. How are native-speaker references used in the TOEPAS 1.0 feedback reports written by raters?

     In the 400 written feedback reports, a frequent recurrence of references to nativeness were found (total=190; native=128; non-native=62). The higher the level, the more references to the native norm per report (e.g., level 5 reports have 3-6 references; level 3, 0-1). References to "nativeness" were found mostly with relation to description of pronunciation. Table 2 presents in detail the uses of "native" and "non-native" in the TOEPAS written reports across skills and proficiency levels.

	native			non-native		
	reports	per report	per skill	reports	per report	per skill
Level 5 (86%)	n=18	6 (n=2) 5 (n=4) 4 (n=5) 3 (n=7)	fluency (n=18) pronunciation (n=14) grammar (n=6) vocabulary (n=19) interaction (n=15) **total (n=72)**	n=5	1 (n=5)	pronunciation (n=5) **total (n=5)**
Level 4 (31%)	n=30	1 Eng. (n=13) 1 L1 (n=14)	fluency (n=6) pronunciation (n=4+16) grammar (n=11) vocabulary (n=3) interaction (n=5) **total (n=45)**	n=29	2 (n=7) 1 (n=22)	pronunciation (n=16) **total (n=16)**
Level 3 (22%)	n=11	1 L1 (n=11)	pronunciation (n=11) **total (n=11)**	n=27	2 (n=13) 1 (n=14)	pronunciation (n=40) **total (n=40)**
Level 2 (9%)	0	0	0	n=1	1 (n=1)	pronunciation (n=1) **total (n=1)**

Table 2. Native-speaker references by skill and proficiency level

As Table 2 shows, the highest number of references to the English native-speaker norms across all sub-skills (fluency, pronunciation, grammar, vocabulary, and interaction) was found in the reports of the highest scalar level (n=72). A large percentage of reports at level 5 (86 %) contained three or more references to native-speaker norms, unlike levels 3 and 4, where that percentage was lower, i.e., 22% and 31% respectively. Although it was reasonable to expect references to native-speaker norms at level 5 because the only descriptor for that level is "Equivalent to a highly articulate, well-educated native speaker," the frequency of these references within each report (3-6) seemed exaggerated. The native-speaker references to pronunciation in levels 3 and 4 were mostly related to lecturers' L1 rather than English. Statements like, "[the lecturer's] pronunciation is marked by native language (Danish) features," were common especially in level 3 reports. The linguistic environment in which the term "native" occured in the report is shown in Figure 1. The larger the size of the word in the diagram, the stronger association with the term "native." As can be seen from the diagram, in addition to the strong association with "articulate, well-educated native speaker" and lecturers' L1s (e.g., Danish, German, Polish), phrases like "near-native accent" and "almost native accent" were also common.
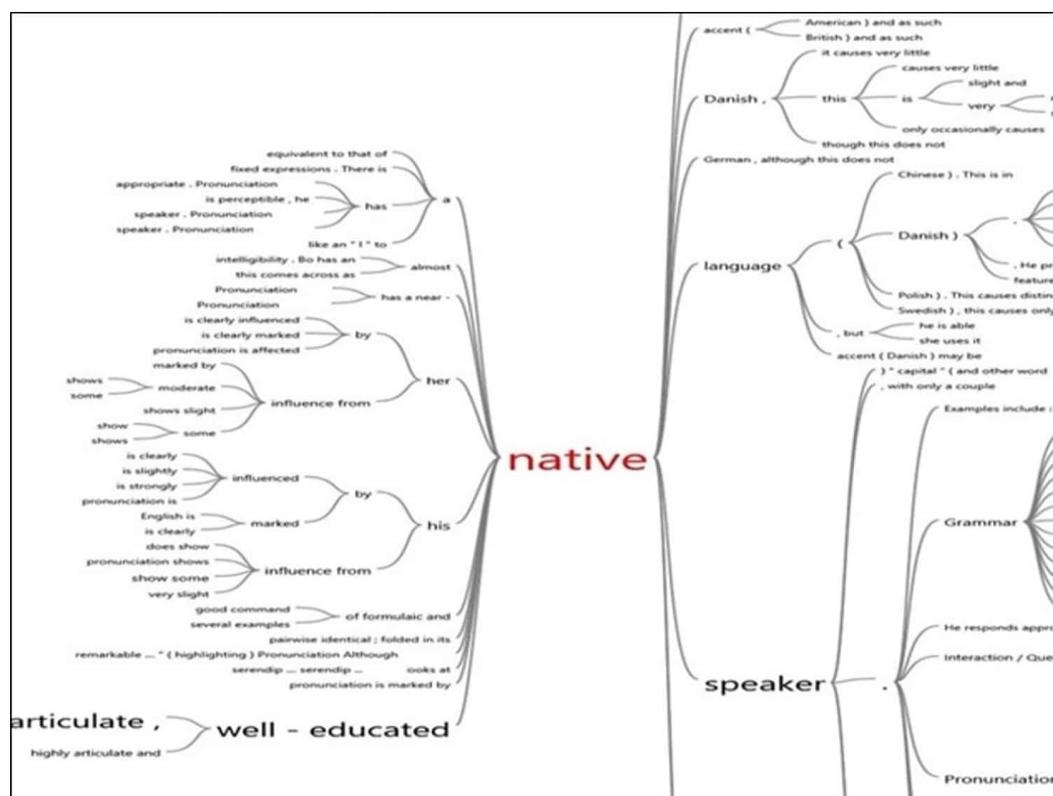
Fig. 1. Uses of the descriptor "native"

While references to "nativeness" occurred across all sub-skills in levels 4 and 5, references to "non-nativeness" were linked only to pronunciation and accent. Level 2, which is below the cutscore for certification, lacked such references. As Figure 2 indicates, the text search query for the term "non-native" in NVivo, resulted in a diagram where "non-native" was most frequently used with terms like, "accent," "features of production," "influences," and "speaker." The diagram also shows that the level of markedness and perceptibility of "mother tongue" influences and "intelligibility" were quite commonly co-occurred with "non-native".
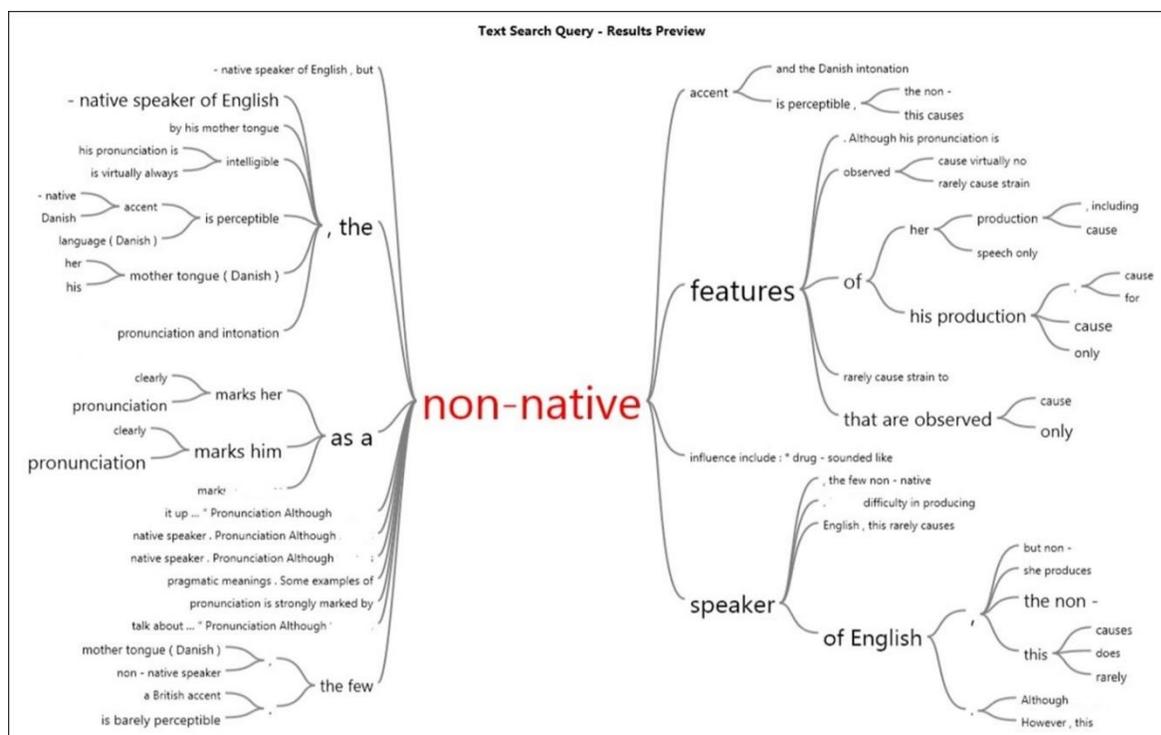
Fig. 2. Uses of the descriptor "non-native"

To gain an improved understanding of the context in which the native-speaker reference was used, following are a few examples of TOEPAS written reports. Rather than abstract references to "the native speaker," to describe the performances, the raters referred to a specific native-speaker norm, in the following cases (examples 1 and 2) American:

> 1) X has a native-like (American) English accent and consistently uses intonation well to convey a range of pragmatic meanings and emphasize main points. He produces all phonological contrasts with accuracy and places stress correctly. In other words, his pronunciation is equivalent to that of an articulate, well-educated native speaker. (Level 5)

2) Although X's L1 is perceptible, he has a native-like (American) English accent and consistently uses intonation well to convey a range of pragmatic meanings and emphasize main points. (Level 5)

X displays an extremely broad range of general, academic, and domain-specific vocabulary as well as fixed expressions. There is a native-like naturalness to his speech as in the following examples… (Level 5)

The following are examples (3 and 4) from levels 3 and 4 performance descriptions, where the raters refer to performances marked by lecturers' L1 or native-like uses of formulaic expressions:

3) Although X's pronunciation and intonation are both marked by his mother tongue (Danish), the non-native features of his production only rarely cause strain to the listener. He produces most sounds with fairly good accuracy, although some contrast errors observed were… (Level 3)

4) X displays a wide range of general, academic and domain-specific vocabulary for effective communication. His language is quite sophisticated in general, and there are several examples of formulaic and native-like expressions. (Level 4)

5.2. How do EMI lecturers perceive their own language proficiency and the use of English in the EMI classroom after receiving the feedback?

Although the interview questions focused on lecturers' perceptions of classroom communication and their own language proficiency, rather than opinions and attitudes towards language norms, references to norms emerged in lecturers' responses. The analysis of lecturers'

responses revealed tensions between the native-speaker ideologies and the EMI classroom realities. Lecturers' reflections on language proficiency seemed to be greatly influenced by the native speaker ideology, while their discourse related to EMI classroom communication acknowledged the fluidity of norms and linguistic resources.

The acceptance of the native-speaker ideologies were evident when the lecturers found the native-speaker reference to have an added value to the top level of the scale. When discussing the TOEPAS feedback reports, some explicitly took pride in receiving the top TOEPAS score because it meant that they sounded like educated native speakers, which they considered an ultimate language learning achievement.

I was certified as a native speaker. I thought that was brilliant. (P05, Law)

You gave me a score of 5, which is the top one … I had English at the level of an articulate native speaker, which is always a nice compliment given that I'm not a native speaker haha, right? (P06, Computer Science)

Those who received lower scores on the TOEPAS idealized the native-speaker model, especially in relation to pronunciation and grammar. Aware of their non-native accent and pronunciation mistakes, the lecturers admitted their lack of expectation to receive the highest TOEPAS score.

I don't know I think I would need courses just to improve my pronunciation … Pronunciation is very important to have a native speaker or somebody who is really, really good at it. (P13, Food and Resource Economics)

I am not a native speaker, and I make some mistakes…That's what I thought and what the certification stated. If I failed, or if it said that I'm equal to a native speaker, then I would have been a little surprised. Of course, I know I'm not like a native speaker. I have an accent and I make grammar mistakes. The verdict was where I expected it to be. (P23, Biomedical Science)

However, when asked to discuss their teaching experiences and communication practices in EMI, the lecturers spontaneously, without any prompts, referred to issues related to the use of ELF in academia context. The lecturers recognized the ELF classroom context and claimed that the multilingual backgrounds of students did not impede communication,

They come from all over the world…The majority comes from … different European countries. Mainly we don't have many native English speakers. Most of them come from … Southern Europe, Spain, Italy, Germany, and some from the Nordic countries…In that sense you know you hear a lot of funny English, but usually that's it's not a problem for communication…that there are lots of alternative ways of pronunciation. (P02, Plant and Environmental Science)

Although many lecturers referred to the native-speaker norm as the highest goal of achievement, with regard to the teaching context, they raised concerns about having the "educated native-speaker" reference in the scalar descriptions and discussed possible speech accommodations (e.g., slow speech rate and simple vocabulary) and grammar adjustments to improve student comprehension levels. For instance, one of the lecturers believed that some students might experience difficulties when listening to native-like speech,

… it can be an advantage sometimes if you don't speak English too well because you could use the sounds more clearly and simple words that are easier for others…that could be part of the English test actually, or it could be at least considered as part of it. (P09, Department of Resource Economic and Food Policy)

Speech accommodation with the purpose to improve students' comprehension levels, i.e. using common vocabulary and rich explanations, was also mentioned as common practice,

If I use fancy words, I have to explain them in other ways. To some I might be too low. The diversity in the student groups language-wise here is enormous. Students from Turkey or Asia are having a hard time. I'm trying to explain things in different ways which may bore those who understood me the first time. (P12, Food Science)

One lecturer even offered an example of grammar innovation that she utilized with some groups of students although the concern that this practice might affect her English proficiency was still present:

Because I've always been working in developing countries I have developed a simplified English…my experience shows that if I don't … If I use 'the' correctly then it's often confusing. They wouldn't understand me. The people from Africa, for instance, they would understand me better if I skip "the" and just take the word and say "a". And, that has affected my English, I'm sure… (P09, Department of Resource Economic and Food Policy)

The lecturers also weighed the importance of being confident versus being accurate, especially in relation to an inner-circle native norm:

…it's really important with speaking and grammar and also the way we talk in English

that we pronounce in the right way. That's important, but it's also important for us to feel

kind of confident, and it's ok that we have our own kind of dialect and that we are not

looking to get perfect in speaking a certain kind of English… (P14, Biomedical Science)

6. Discussion

Two main findings characterize the present study: influence of native-speaker ideologies

in TOEPAS feedback reports and lecturers' conflicting views regarding the correctness of native-

like language and the acceptance of variation in the EMI classroom. The "native speaker"

reference was reiterated with clear dominance in the feedback reports at the highest level, which

was expected given its presence in the scalar descriptors. A less expected finding was that the

emphasis of the feedback reports rested predominantly on the structural linguistic characteristics.

The feedback seemed to promote lexical and grammatical accuracy and native-like accent as

important aspects of EMI lecturers' English language production. The compartmentalized view

of language systems transpired in lecturers' responses because most lecturers compared

themselves against the "native speaker" norm when they discussed their proficiency levels,

despite the lack of any direct references to norms in the interview questions. The lecturers

highlighted their "deficiencies" with regard to pronunciation and grammar rather than pragmatics

when they described their own speaking skills. Although the lecturers' self-perceptions may have

been guided by their exposure to native-speaker ideologies in their language education, these

have been obviously confirmed in the feedback reports.

Concomitantly, the lecturers exhibited acute awareness of the ELF strategies of the EMI

context when they discussed the linguistic and cultural heterogeneity of the student population.

Therefore, lecturers emphasized different points when they talked about their own language proficiency than when they discussed classroom communication. While lectures tended to emphasize pronunciation and grammatical accuracy in relation to their own speech production, they underlined confidence development and ability to accommodate when they considered communication with students. In other words, some of the lecturers were more aware of the multilingual context in which they found themselves, and they explicitly distinguished between a competent and a native speaker, which corroborated findings from other EMI contexts (Pilkinton-Pihko, 2013). Such awareness and distinction is clearly beneficial in EMI programmes where the norms are emergent (Söderlundh, 2013).

As mentioned earlier, the study's purpose was to investigate whether the TOEPAS functioned as envisioned and what revisions, if any, had to be undertaken. The TOEPAS was designed to assess whether lecturers possess adequate oral English proficiency for the EMI classroom. For that reason, the TOEPAS feedback report was intended to raise awareness about the strengths and weaknesses of lecturers' communication in EMI. However, the findings suggested that due to its strong focus on accuracy, the TOEPAS feedback report had failed to reflect the communicative realities of the EMI classroom and seemed to contradict the purpose of the TOEPAS assessment and feedback. The linguistic categories assessed and reported in the TOEPAS written feedback needed an improved alignment with the actual target language use (TLU) in the EMI domain in order to warrant assessment validity, as Shohamy (2017) recommends in her work. If the formative feedback report was to raise awareness among lecturers about how to communicate effectively in the multilingual EMI classroom where English is used as a lingua franca then it seems more reasonable that it focuses on the local and functional rather than the normative and structural aspects of language.

6.1. New version of the test: TOEPAS 2.0

The findings from the study informed the revision of the TOEPAS scale and the formative feedback format. In addition to the removal of the native speaker reference from the scale, some of the scalar categories were revised to reflect the pragmatic strategies lecturers use when they teach. Based on the analysis of the video-recordings with performances at different proficiency levels and previous research findings regarding pragmatically effective speakers in the EMI classroom (Björkman, 2010; Suviniitty, 2012), new scalar categories were incorporated. These include descriptors related to lecturers' abilities to present content material effectively using summaries, examples, or emphasis of important points, as well as their abilities to deliver their lectures in an organized and structured manner exploiting different types of explicit and implicit signposting elements. Some structural linguistic elements (e.g., grammar, syntax) remain among the descriptors, but their prominence was reduced.

The results were particularly constructive in the revision of the TOEPAS written feedback format. Instead of compartmentalized description of the different linguistic aspects of the lecturer's performance, the new feedback report raises awareness about the pragmatic functions of the different language characteristics observed in the lecturer's performance on the test and offers recommendations on various communicative strategies for the EMI classroom. Since the purpose of the TOEPAS is to make inferences about lecturers' ability to deal with the communicative demands of the EMI classroom, then discussing the test results with reference to the TLU domain for which the inferences are made seems reasonable. In other words, although the test measures English language proficiency, the results become more meaningful for the lecturers if they are framed within the multilingual context of EMI programmes.

7. Conclusions

       The growing mobility of students and lecturers and the linguistic heterogeneity of the local contexts have resulted in increasingly multilingual universities. Given these linguistic trends, renegotiation of language policies at universities may need to be established in order to acknowledge the role of multilingualism in teaching and learning in academia. Understanding the role of English in the multilingual EMI classroom is also important because of both student and teacher diversity. However, even if policies for multilingualism are reached, the translation of these policies into concrete practices may seem challenging. This challenge seems quite pronounced in language testing situations, in which a heightened tension exists between the need for standardization and the observed variation of norms.

       The current chapter discussed such a tension between lecturers' reported classroom practices and the norms promoted in the written feedback reports of the TOEPAS, which is an oral English proficiency test. Based on lecturing test performance analysis and interviews with lecturers, the solution for this tension was the development of (1) a rating scale with focus on the language functions and the pragmatics of English in the university classroom, rather than the accuracy of phonology and lexicogrammar, and (2) a formative feedback procedure that contextualizes lecturers' language uses in the teaching situation, keeping in mind the heterogeneous student population.

References

Bachman, L., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford: Oxford University Press.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world.* Oxford: Oxford University Press

Baker, B. (2017). *Expanding the construct of 'passive bilingualism' in a listening test for university professors: Incorporating translanguaging.* Paper presented at the 2017 American Association of Applied Linguistics (AAAL) Conference, Portland, OR.

Ball, P., & Lindsay, D. (2013). Language demands and support for English-medium instruction in tertiary education. Learning from a specific context. In A. Doiz, J. M. Sierra, & D. Lasagabaster (Eds.), *English-medium instruction at universities: Global challenges* (pp. 44-64). Bristol, UK: Multilingual Matters.

Björkman, B. (2010). So you think you can ELF: English as a lingua franca as the medium of instruction. *Hermes–Journal of Language and Communication Studies, 45,* 77-96.

Björkman, B. (2011). Pragmatic strategies in English as an academic lingua franca: Ways of achieving communicative effectiveness? *Journal of Pragmatics, 43*, 950-964.

Borg, S. (2009). *Introducing language teacher cognition.* Retrieved 9 April 2019 from http://www.education.leeds.ac.uk/research/files/145.pdf.

Canagarajah, S. (2016). Translingual writing and teacher development in composition. *College English, 78*(3), 266–273.

Chopin, K. (2014). Reconceptualizing norms for language testing: Assessing Engish language proficiency from within an ELF framework. In Y. Bayyurt & A. Sumru (Eds.), *Current perspectives on pedagogy for English as a lingua franca.* Berlin: De Gruyter Mouton.

Coleman, J., Hultgren, K., Li, W., Tsui, C. F. C., & Shaw, P. (2018). Forum on English-medium Instruction. *TESOL Quarterly, 52*(3), 701-720.

Davies, A., Hamp-Lyons, L., & Kemp, C. (2003). Whose norms? International proficiency tests in English. *World Englishes, 22* (4), 571-584.

Dimova, S. (2017a). Pronunciation assessment in the context of World Englishes. In O..Kang & A. Ginther (Eds.). *Assessment in second language pronunciation* (pp. 49-66). Oxfordshire, UK: Routledge.

Dimova, S. (2017b). Life after oral English certification: The consequences of the Test of Oral English Proficiency for Academic Staff for EMI lecturers. *English for Specific Purposes, 46*, 45-58.

Dimova, S., & Kling, J. (2018). Assessing English-medium instruction lecturer language proficiency across disciplines. *TESOL Quarterly, 52*(3), 634-656.

Douglas, D. (2000). *Assessing Language for Specific Purposes.* Cambridge Language Assessment. Cambridge: Cambridge University Press.

Elder, C., & Harding, L. (2008). Language Testing and English as an International Language Constraints and Contributions. *Australian Review of Applied Linguistics, 31*(3), 34-1.

García, O., & Li, L. (2014). Translanguaging and education. In *Translanguaging: Language, bilingualism and education* (pp. 63-77). London: Palgrave Macmillan.

Harding, L., & McNamara, T. (2017). Language assessment: The challenge of ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (pp. 570-582). London: Routledge.

Hellekjær, G. O. (2010). Lecture comprehension in English-medium higher education. *Hermes–Journal of Language and Communication Studies, 45*, 11-34.

Hynninen, Niina (2011) The practice of 'mediation' in English as a lingua franca interaction. *Journal of Pragmatics*, 43 (4), 965–977. DOI: 10.1016/j.pragma.2010.07.034.

Jenkins, J., & Leung, C. (2017). Assessing English as a lingua franca. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment (3rd ed.)* (pp. 103–117). New York, NY: Springer.

Joe, Y. J., & Lee, H.-K. (2013). Does English-medium instruction benefit students in EFL contexts? A case study of medical students in Korea. *Asia-Pacific Educational Research, 22*(2), 201–207. https://doi.org/10.1007/s40299-012-0003-7

Julien, H. (2008). Content analysis. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods*, *2*. Thousand Oaks, CA: SAGE Publications.

Klaassen, R. G., & Bos, M. (2010). English language screening for scientific staff at Delft University of Technology. *Hermes–Journal of Language and Communication Studies, 45,* 61-75.

Kling, J., & Hjulmand, L.-L. (2008). PLATE–Project in language assessment for teaching in English. In R. Wilkinson, & V. Zegars (Eds.), *Realizing content and language integration in higher education* (pp. 191-200). Maastricht: Universitaire Pers Maastricht.Kling, J. (2016). *"I don't teach language": The NNS Lecturer and EMI.* Paper presented at TESOL Italy. http://tesolitaly.org/new/wp-content/uploads/2016/11/TESOL-Italy-plenary-Joyce-Kling.pdf

Kling, J., & Stæhr, L. S. (2011). Assessment and assistance: Developing university lecturers'
skills through certification feedback. In R. Cancino, L. Dam, & K. Jæger (Eds.), *Policies,
principles, practices: New directions in foreign language education in the era of
educational globalization* (pp. 213-245). Newcastle upon Tyne, UK: Cambridge Scholars
Press.

Kling, J. & Stæhr, L. S. (2012). *The development of the Test of Oral English Proficiency for
Academic Staff (TOEPAS).*

http://cip.ku.dk/forskning/cip_publikationer/CIP_TOPEPAS_Technical_Report.Pdf

Li, L. (2011). Moment analysis and translanguaging space: Discursive construction of identities
by multilingual Chinese youth in Britain. *Journal of Pragmatics, 43*(5), 1222-1235.

Lowenberg, P. (1993). Issues of validity in test of English as a world language: whose standards?
World Englishes, 12 (1), 95-106.

Macaro, E. (2018). *English medium instruction.* Oxford: Oxford University Press.

Mason, J. (2004). *Semistructured interview. The SAGE encyclopedia of social science research
methods.* Thousand Oaks, CA: Sage Publications.

Mauranen, A. (2007). Hybrid voices: English as the lingua franca of academics. In K. Flottum,
T. Dahl and T. Kinn (Eds.) *Language and Discipline Perspectives on Academic
Discourse*. Cambridge: Cambridge Scholars Press. 244–259.

Mauranen, A. (2010). Features of English as a lingua franca in academia. *Helsinki English
Studies, 6,* 6-28.

Mauranen, A. (2012). *Exploring ELF: Academic English Shaped by Non-Native Speakers*. Cambridge: Cambridge University Press.

Mauranen, A., Hynninen, N., & Ranta, E. (2010). English as an academic lingua franca: The ELFA project. *English for Specific Purposes, 29*(3), 183-190.

Menken, K., & Shohamy, E. (2015). Invited colloquium on negotiating the complexities of multilingual assessment, AAAL Conference 2014. *Language Teaching, 48*(3), 421-425.

Mortensen, J. (2014). Language policy from below: Language choice in student project groups in a multilingual university setting. *Journal of multilingual and multicultural development, 35*(4), 425-442.

Murata, K., & Iino, M. (2017). In J. Henkins (ed.), *EMI in higher education: An ELF perspective. In The Routledge handbook of English as a lingua franca,* (pp. 400-412). London: Routledge.

Newbold, D. (2014). Engaging with ELF in an entrance test for European university students. In Y. Bayyurt & A. Sumru (Eds.), *Current perspectives on pedagogy for English as a lingua franca.* Berlin: De Gruyter Mouton.

Pecorari, D., & Malmström, H. (2018). At the Crossroads of TESOL and English Medium Instruction. *TESOL Quarterly, 52*(3), 497-515.

Pilkinton-Pihko, D. (2013). *English-medium instruction: Seeking assessment criteria for spoken professional English. Doctoral dissertation (monograph),* University of Helsinki. HELDA E-thesis: http://urn.fi/URN:ISBN:978-952-10-9520-7

Ranta, E. (2006). The 'attractive' progressive – Why use the -ing form in English as a lingua franca? *Nordic Journal of English Studies 5* (2), 95–116.

Seidman, I. (2013*). Interviewing as qualitative research: A guide for researchers in education and the social sciences.* New York: Teachers College Press.

Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal, 95*(3), 418-429.

Shohamy, E. (2017). ELF and critical language testing. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (pp. 583-593). London: Routledge.

Smit, U. (2010). *English as a Lingua Franca in Higher Education: A Longitudinal Study of Classroom Discourse*. Berlin: De Gruyter Mouton.

Suviniitty, J. (2012). *Lectures in English as a Lingua Franca: Interactional Features. Doctoral dissertation (monograph).* University of Helsinki. HELDA E-thesis: http://urn.fi/URN:ISBN:978-952-10-8540-6.

Söderlundh H. (2013) Language Choice and Linguistic Variation in Classes Nominally Taught in English. In: Haberland H., Lønsmann D., Preisler B. (eds) *Language Alternation, Language Choice and Language Encounter in International Tertiary Education. Multilingual Education, vol 5*. Dordrecht: Springer.

University of Freiburg, L. T. C. (2016). *English medium instruction - Qualitätssicherung*. Retrieved from https://www.sli.unifreiburg.de/englisch/emi/quality?set_language¼e

Wächter, B., & Maiworm, F. (Eds.). (2014). *English-taught programmes in European higher education: The state of play in 2014*. Bonn: Lemmens Medien GmbH.

Yeh, C. C. (2012). Instructors' perspectives on English-medium instruction in Taiwanese universities. *Curriculum and Instruction Quarterly, 16,* 209–232.