THE ROLE OF FEEDBACK IN THE DESIGN
OF A TESTING MODEL FOR SOCIAL JUSTICE

Slobodanka Dimova
University of Copenhagen
slobodanka.dimova@hum.ku.dk

Internationalization of higher education has resulted in an increased establishment of English medium instruction (EMI) courses at non-Anglophone universities. Due to the growing concerns about non-native English speaking lecturers' ability to teach in English, universities have started implementing policies enforcing assessment of EMI lecturers' English proficiency. However, research on the effectiveness and the social consequences of these assessments remains limited, especially in terms of the power imbalances such assessments may create at the university workplace. Based on the principle that the oral English proficiency certification should provide benefits for, instead of simple exclusion of, test-takers who have less-advantaged starting position, i.e., lecturers with lower English proficiency (Davies 2010), the article argues that the language assessment models for EMI certification should put emphasis on the formative feedback, and, thereafter, proposes what may be considered relevant and effective feedback content. The discussion in this article is based on the results from a mixed-methods study that examines the utility of the language-related feedback for test-takers, at two non-Anglophone universities, who took a performance-based oral English proficiency test for EMI certification. Three main data collection procedures were undertaken: reports with written feedback and two surveys. Results suggest that the test-takers found the aspects of the feedback that grounded their English language uses in the EMI domain useful and appreciated the opportunity to discuss their challenges and uncertainties during oral feedback session.

Keywords: language tests, EMI, feedback, social justice, EMI certification

ПОВРАТНИТЕ ИНФОРМАЦИИ
КАКО ОСНОВА ЗА СОЦИЈАЛНА ПРАВДА
ПРИ ЈАЗИЧНОТО ТЕСТИРАЊЕ

Слободанка Димова
Универзитет во Копенхаген
slobodanka.dimova@hum.ku.dk

Интернационализацијата на високото образование доведе до пораст на бројот на предмети од различни научни области што се предаваат на англиски јазик и тоа на универзитетите во земјите каде што англискиот не се говори како мајчин јазик. Со воведувањето на ваквите предмети се зголеми и бројот на мерки (тестови, сертификати) што универзитетите ги преземаат за да ја одредат способноста на професорите да предаваат на англиски јазик, иако англискиот не им е мајчин. Но, дали резултатите од тестирањето се ефикасни и дали предизвикуваат социјална нерамнотежа на универзитетот како работно место, ретко претставува предмет на научните истражувања. Врз основа на принципот дека тестовите за познавање на англискиот јазик треба да претставуваат средства за помош, а не ексклузија, на професорите што имаат потреба од поддршка (Davies 2010), т.е. оние што имаат лимитирани способности во употребата на англискиот јазик, важно е да се стави акцент на повратните информации што професорите ги добиваат врз основа на тестот. Во статијава се презентираат резултатите од истражување со мешана (квантитативна и квалитативна) методологија, чија цел беше анализа на ефикасноста на повратните информации од тестот по говорен англиски јазик за сертификација на универзитетски професори што предаваат на англиски јазик. Собраните податоци вклучуваат извештаи со повратни информации и анкетни податоци. Резултатите упатуваат на тоа дека, ако се контекстуализирани повратните информации, тие се ефикасни, бидејќи им овозможуваат на тестираните професори да се запознаат со стратегиите потребни за успешна комуникација во училницата

Клучни зборови: јазични тестови, англискиот како наставен јазик, повратни информации, социјална правда, сертификација на универзитетски професори

1 Introduction

The past two decades have marked a steep growth of English medium instruction (EMI) programs at non-Anglophone universities. This growth has been governed by various factors at the global, international, national, and local levels (Hultgren, Jensen and Dimova 2015). At the global level, the General Agreement on Trades and Services from 1995 has turned higher education (HE) into a commodity and promoted university competition for the HE market (Philipson 2015). Given the high recruitment of international students and lecturers at North American, British, and Australian universities, non-Anglophone universities could become competitive in the HE market by offering EMI courses and programs. At the European level, on the other hand, Englishization has been fueled by the Bologna declaration (1999), which promotes students and lecturers:

> …mobility by overcoming obstacles to the effective exercise of free movement with particular attention to:
> • for students, access to study and training opportunities and to related services
> • for teachers, researchers and administrative staff, recognition and valorisation of periods spent in a European context researching, teaching and training, without prejudicing their statutory rights (Bologna Declaration 1999: n.p.).

National strategies for internationalization are often reflected in the local university policies that emphasize internationalization of courses, programs, and degrees. The University of Copenhagen (UCPH) strategy from 2008, for example, highlights the use of English, in addition to Danish, in order to recruit international and retain the best domestic students (University of Copenhagen 2008).

The implementation of EMI at non-Anglophone universities has impacted teaching and learning in HE. One of the obvious consequences is that lecturers teach and students learn in a foreign language (FL). Therefore, a number of heated debates have occurred regarding the quality of teaching and learning in these EMI situations, where most stakeholders are FL speakers of English, and concerns have been raised about lecturers' identities (Wilkinson and Zegers 2006). Although sometimes EMI lecturers may be expected to step in as language teachers because of the varied English proficiency among the students, many of them reject the role of language teachers insisting that their focus should remain on the disciplinary content (Airey 2012). Moving to EMI has also led to questions about lecturers' professional, teacher, and linguistic identities. While experienced teachers may feel confident with their teacher identity in EMI, many remain insecure about their linguistic identity due their challenges with language (Kling 2013). These changes in the classroom context have led to many controversies regarding the implementation of EMI that have reached debates a political level (Blattès 2018, Campagna 2017, Jensen and Thøgersen 2011, Pulcini and Campagna 2015, Santulli 2015).

Due to these debates and controversies, university managements have begun to scrutinize lecturers' English proficiency levels in order to ensure that they can use English effectively in the EMI classroom. Some universities require standardized

English test scores to certify EMI lecturer English proficiency (e.g., International English Language Testing System (IELTS), Test of English as a Foreign Language (TOEFL) or other CEFR-based assessment), others have developed local certification procedures (Kling and Hjulmand 2008, Ball and Lindsay 2013, Kling and Stæhr 2011).

Regardless of whether universities require standardized or locally-developed tests, the English language requirements have caused added pressure on EMI lecturers. Not only do lecturers have to switch to English, their teaching in English is also scrutinized (Dimova 2017). In such tense situations, questions about the ethics, fairness, and justice of language assessment arise and need to be negotiated.

Research on the effectiveness and the social consequences of English language assessments for EMI lecturers remains limited, especially in terms of the power imbalances such assessments may create at the university workplace. Based on the principle of ethics that the oral English proficiency certification should provide benefits for, instead of simple exclusion of, test-takers who have a less-advantaged starting position, i.e. lecturers with lower English proficiency (Davies 2010), this article will argue that the language assessment models for EMI certification should put emphasis on relevant formative feedback.

Ethical considerations, such as fairness and justice, based on Kant's principles of dignity and reason, commonly used by Kantian philosophers, are applicable to language assessment in that morally acceptable assessment practices should not subject test-takers to humiliating exam conditions that may affect their dignity (Deygers 2019). Rawls (2001), a follower of Kantian ideas, defines justice through fairness and absence of bias and proposes two main principles of justice. The first principle postulates that a society is just only if it provides equal access to the basic liberties. The second principle is the precondition for the first one, and it requires that equal access to the same positions should still be available to all citizens when inequalities arise, and if inequalities do arise, they should favor citizens with less advantaged positions.

Rawls has been quite influential in the theories of language assessment. Messick's (1989) conceptualization of fairness is related to bias avoidance and evidence of lack of construct-irrelevant variance, while he associates justice with the societal and individual consequences a test incurs. Assessment results should represent the test-takers' levels of a particular language ability that is theoretically constructed. Therefore, the variation of scores among test-takers should represent the variation of ability levels among the test-takers. If the sore variance is construct-irrelevant, i.e. if the differences in scores come from abilities other than the measured ability, then the test is unfair and biased. Given that assessment results are used for decision-making, they have consequences for both test-takers and test users. For example, content teachers in EMI settings are allowed or not to teach in English based on their speaking abilities represented by English test scores. If these scores are based on assessments that represent other constructs (e.g., grammar tests), then the results will be irrelevant and the test consequences, i.e. the decisions made based on these results, will not be just both for the lecturers and the students because the selection of lecturers to teach in EMI will not be appropriate.

McNamara and Ryan (2011) further clarify the distinction between fairness and justice in language assessment in that fairness relates to establishment of equality, equity, and bias avoidance, and justice depends on the defensibility of the values that the test represents. Therefore, fairness is assessment-internal and can be evaluated through the interpretation and use of the psychometric properties of assessments, while justice is assessment-external and depends on the social values and the stakes the assessment has in the policy context. In that sense, although fairness in many cases is a precondition for just assessments, it is not the sole condition for justice; even if an assessment lacks bias, it cannot be just if it is founded on unjust values.

Implementing EMI across university courses and programs at non-Anglophone universities may not be a just policy if the university content lecturers are suddenly required to switch to teaching in English without provision of language support or other incentives, like pedagogical training, extra preparation time, or funding. Moreover, language assessments that are instruments of such policies may create unjust social consequences for content lecturers because these assessments will create unequal access to social positions for different lecturer groups in the departments. Some content lectures with extensive teaching experience, unique expertise in their research disciplines, and shared first language with students may acquire a lower status in their departments if they obtain lower English proficiency results obtained on the assessment, while inexperienced content lecturers may obtain important teaching assignments because of their high English proficiency. Therefore, even if a language assessment for EMI lecturers who are non-native speakers of English meets all the assessment-internal psychometric requirements for fairness and bias avoidance, this assessment may still lack the assessment-external requirements for justice because the underlining social values that the assessment policy imposes are unjust.

One way to meet the assessment-external requirements for justice in language assessments for EMI lecturer certification is by designing an assessment instrument with emphasis on its formative and supportive purpose rather than its selective role, i.e. dividing lecturers in certified and non-certified groups. Results from external language tests have traditionally been reported as scores, while formative feedback has been primarily used in instructional settings. Classroom feedback tends to be conceptualized as the teacher's evaluation of the learner's performance on a specific task or a summary of the learner's performance over a certain instructional period. Shute emphasizes the process of learning by defining formative feedback as, "information communicated to the learner that is intended to modify the learner's thinking or behavior for the purpose of improving learning" (2008: 1). The relationship between instruction and feedback can be represented on a continuum where these two activities are separate and distinct at one end and completely integrated at the other (Hattie and Timperley 2007). The meta-analyses of the numerous studies in education have yielded differentiated and conflicting results regarding the characteristics of effective feedback (Hattie and Timperley 2007, Shute 2008). Clear and specific directions for improvement may be considered more effective than general feedback statements. However,

specificity and detail can lead to complexity, which may impose processing difficulties and have debilitating effects.

In the absence of established training programs available to EMI lecturers, the assessment method can become an opportunity to integrate assessment, feedback, and learning by providing extensive formative feedback and information about available resources for support. The formative role of the feedback from the test is warranted only if the feedback is relevant and informative in relation to the communicative domain of language use, in this case the EMI classroom. In order to exemplify the formative role of language assessment for EMI lecturers, the present study examines the features and the effectiveness of the feedback from the Test of Oral English Proficiency for Academic Staff (TOEPAS), which is an oral English proficiency test used for EMI teacher certification at the University of Copenhagen.

2 The research context: The Test of Oral English Proficiency for Academic Staff (TOEPAS)

In 2008, the management at the University of Copenhagen commissioned a test for certification of EMI lecturers in order to ensure that the lecturers have adequate oral English proficiency to teach in the EMI classroom. In the negotiations about the format of the certification procedure, the Danish researcher and teacher unions demanded that all measures be taken to avoid any negative consequences for the lecturers, i.e. the certification does not become just a tool for power imbalance and exclusion but to offer equal opportunity for lecturers to learn (Kling and Stæhr 2011).

For the purpose of equality for all test-takers, the Test of Oral English Proficiency for Academic Staff (TOEPAS) was designed in the format of a simulated lecture. This format allowed for eliciting classroom-related language with the use of standardized test administration and rating procedures (Kling and Stæhr 2011). To avoid inequality among university lecturers, pedagogical competences were not included in the test construct, given that the pedagogical competence of lecturers who teach in their first language (Danish) were not evaluated. The test criteria were designed to focus on how lecturers use language in the teaching domain rather than how they teach the disciplinary content. The performances from the simulated lecture are rated live by two raters who are also EMI experts and have been trained to use the criteria consistently. Test-taker equity, on the other hand, was established by including a video recording and detailed written feedback regarding test-taker's (lecturer's) individual performance together with the test score in the result report (for more information about the format and the psychometric properties of the test, see Dimova and Kling 2015, Dimova and Kling 2018, Kling and Dimova 2015). In other words, the emphasis rested on the formative aspects of TOEPAS rather than its screening purpose.

In the first version, the written TOEPAS feedback described the test-takers' performance regarding fluency, pronunciation, grammar, vocabulary, and interaction. An important characteristic of the report was inclusion of examples and references from the actual performance to support the descriptions (see Dimova 2017, 2020).

Taking into consideration the importance of the TOEPAS formative feedback, the utility of the written feedback report and the consequences for the tested lecturers was examined (Dimova 2017). The results suggested that the most useful aspect of the TOEPAS feedback reports was the examples from lecturers' own performances, but the feedback was not optimized because the lecturers lacked familiarity with technical linguistic terminology included in the descriptions accompanying the examples. Based on these results, Dimova (2017) recommended to redesign the TOEPAS feedback so that lecturers could be informed about the functions and the uses of the feedback information for the purpose of maximizing the feedback effects and minimizing misuses.

Based on these recommendations, the TOEPAS feedback procedure was revised, so now the written report includes reduced use of technical linguistic terms, discusses the performance descriptions from a functional perspective (i.e., the function of language in EMI classroom communication), and integrates recommendations for improvement. An oral feedback session provided by the raters was also incorporated in order to provide opportunities for clarifications of the written report and for drawing lecturers' attention to the feedback functions and its uses. After its revision, the TOEPAS was implemented at an additional university in Denmark to substitute their previous certification procedure.

3 Methodology

In order to find out whether the revised version of the TOEPAS feedback provides opportunities for lecturers to learn, the following research questions were developed:

1. What are the characteristics of the TOEPAS feedback?
2. What is lecturers' perception of TOEPAS feedback effectiveness?

A convergent parallel mixed-method design (Creswell and Clark 2007) was applied to investigate the research questions. Data collection included feedback reports and two online survey questionnaires. The convergent mixed-method approach allowed for independent analysis of each data set but collective interpretation of all data types concerning the research questions (Creswell and Clark 2007). The different types of data were thematically coded so that they could be interpreted in a contrastive and a complementary manner in order to reach comprehensive answers to the research questions. The framework for analysis included five variables that have been found to influence the feedback effect: functions, specificity, complexity, timing, and cognitive load.

Functions. Feedback can be directive (refers to concrete parts of task to be improved) or facilitative (overall suggestions for improvement) (Black and William 1998, Underwood 2008).

Specificity. It refers to the level of information provided in the feedback and can range from general (general advice for improvement) to specific (provision of correct answers) (Phye and Sanders 1994).

Complexity. Feedback information can be simple (overall evaluation) or complex (detailed information about performance) (Dempsey, Driscoll and Swindell 1993, Mason and Bruning 2001, Narciss and Huth 2004).

Cognitive load. The specificity and complexity of feedback can lead to different levels of cognitive load, ranging from not demanding to very demanding (Phye and Sanders 1994).

Timing. The feedback that is delivered right after the task performance is considered immediate, while feedback delivered after a certain period of time (hours, days, weeks) is characterized as delayed (Kulhavy and Anderson 1972, Surber and Anderson 1975).

3.1 Data collection and analysis

The feedback procedure and the content of the new version of the TOEPAS feedback were analyzed in terms of the five feedback variables (functions, specificity, complexity, cognitive load, and timing). With the use of NVivo, the word frequency of the feedback reports (n=90) was analyzed in order to identify technical terms.

Two surveys were distributed online to all lecturers (n=90) who were tested in the period between October 2017 and December 2020. The lecturers came from a wide range of disciplinary backgrounds and held different academic positions, ranging from doctoral students to full professors. The first survey was distributed to the lectures immediately after the test session as part of the ongoing evaluation of the testing procedure and included five open-ended questions asking lecturers to evaluate the initial distribution of information about the TOEPAS, the certification session, and the written and the oral feedback, as well as to provide suggestions for improvement. The second survey was administered to obtain specific information about lecturers' perceptions of the TOEPAS that the open-ended questions failed to elicit. This questionnaire included seven multiple- choice questions, three yes/no questions, and four matrix questions with Likert-type items. The multiple-choice questions elicited test-takers' background information, like their TOEPAS score, years of teaching experience, discipline, and language(s) used for instruction. The yes/no questions asked whether the lecturers used the video and the written and oral feedback, while the matrix questions contained evaluative Likert-type items with statements regarding the result reporting procedures and the feedback. The response rate was 51% for the first and 29% for the second survey. The lower response rate for the second survey was expected because it followed the qualitative survey in which many lecturers provided extensive responses.

4 Results and discussion

Data from the written feedback report contributed to answering the research question regarding the characteristics of the revised TOEPAS feedback procedure. Previous literature suggests that effective feedback should include elements of verification (confirming the correctness) and elaboration (guide to correctness) (Mason and Bruning 2001) and that it should reduce uncertainty about task performance (Ashford, Blatt and VandeWalle 2003). Although feedback tends to elaborate on

areas that need improvement, the TOEPAS feedback reports elaborate on the verification of the performance's strengths:

> You clearly progress through your lesson through effective transitions that provide structure, and you use intonation (speech melody) and sentence stress to help the listener understand the organization. Moreover, you emphasize important content and terminology (TOEPAS feedback report A).

As can be seen from the example, the comments are written in present tense, which suggests that the comments are an extrapolation rather than just a description of the observed performance. Tests are used to make assumptions about test-takers' ability to use language in real life. Therefore, if test-takers display the ability to use certain linguistic structures during test tasks, there is an assumption that they are able to use the same structures in contexts that the tasks represent. Therefore, the extrapolation facilitates the transferability of the observed language use to classroom situations beyond the test session, which means that the lecturers who receive the feedback do not interpret it only within the constraints of the test. In terms of feedback functions, this extrapolation suggests that the TOEPAS written feedback is mostly facilitative with some directive elements provided in the recommendation section [e.g., "Our recommendation is to try to expand your general vocabulary range" (TOEPAS feedback report A)]. One of the TOEPAS reports provides a facilitative description that generalizes the area for improvement, so the test-taker can view the video and identify the actual errors:

> You utilize a wide range of grammatical (word) forms and (sentence) structures effectively. The few grammatical mishaps do not impede intelligibility. (TOEPAS feedback report B)

Given that the purpose of the TOEPAS is to provide feedback that is applicable in real life situations, the feedback's facilitative purpose could also be identified through its contextualization in the target language use (TLU) domain (Bachman and Palmer 1996), i.e. the EMI classroom. Most survey respondents agreed that the feedback has this facilitative purpose because of its relevance for the EMI situation (66%), as well as the usability of the feedback (77%) and the recommendations (78%). Figure 1 summarizes the findings.
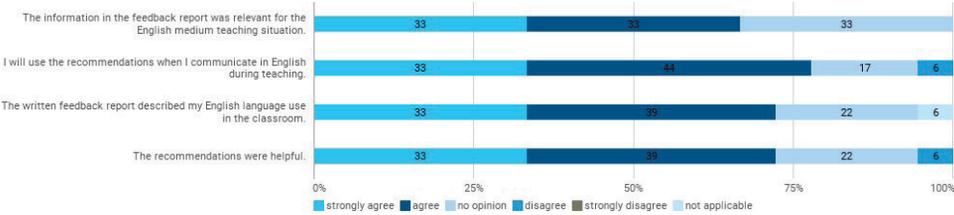

Figure 1. *Perceived usefulness of the TOEPAS feedback report*

Despite the facilitative nature of the TOEPAS feedback, the comments are rather specific because they do focus on particular aspects of the performance. Specificity has been previously associated with effectiveness (Shute 2008). In the TOEPAS feedback report B, the general statement about grammar was supported by specific quotes from the performance and their corrections:

e.g. "bacteria you need to be fast to develop antibiotics to" (for/against)
e.g. "they can be find in the water we drink" (found)
e.g. "I drew a diagram to explain what is the difference" (what the difference is)
(TOEPAS feedback report B).

In other words, the concrete examples taken from the performance increase the specificity of the comments:

The lecture is easy to follow because you use effective transitions to provide structure throughout the performance, including communicative strategies such as paraphrasing and illustration. You clearly highlight key topics:

e.g. "first I'd like to tell you something about the diffusion of the membrane"
(organization of lecture)
e.g. "Today I am going to talk about the active transport" (organization of lecture
–signposting) (TOEPAS feedback report C)

Given the level of specificity and detail, the TOEPAS written feedback remains quite complex, which could be problematic because it can lead to increased cognitive load (Shute 2008). The new version of the TOEPAS feedback includes fewer technical terms and more explanations or definitions of certain concepts. As can be seen in the example from the TOEPAS feedback report A, the term "intonation" was accompanied with "speech melody" as a clarification. The oral feedback session provides the raters with an opportunity to discuss the comments, give further explanations if lecturers needed further clarification, and draw attention to the most important aspects of the feedback.

The word frequency analysis also showed that unlike the first version of the TOEPAS feedback format, which included linguistic terms such as "discourse markers", "sound contrasts", and "pragmatic meaning" (see Dimova 2017), the most common content words in the revised version are related to the context (lecture, questions) (see Table 1).

Word	Count	Weighted Percentage (%)	Similar Words
lecture	332	001	lecture, lecturer, lectures, lecturing
vocabulary	257	001	vocabulary
questions	251	001	question, questions
sounds	172	001	sound, sounded, sounding, sounds
structure	169	001	structure, structured, structures, structuring
organized	154	000	organization, organizations, organize, organized
explain	151	000	explain, explained, explaining
sentence	145	000	sentence, sentences
examples	134	000	example, examples
transitions	119	000	transition, transitioning, transitions
stress	117	000	stress
material	115	000	material, materials
grammatical	111	000	grammatical
intonation	109	000	Intonation
performance	109	000	perform, performance
respond	106	000	respond, responded, responding
hesitations	93	000	hesitant, hesitate, hesitation, hesitations

Table 1. *TOEPAS feedback report: word frequency*

The feedback remains delayed because it cannot be processed immediately after the test performance. However, the delayed feedback is supported with the video recording and the oral feedback. In other words, lecturers can watch the video to relate the feedback to the performance, while the oral feedback helps them to go over the report again and take action if needed.

In terms of effectiveness (RQ2), the adequacy of the feedback was confirmed with the survey. All respondents understood the report and found it informative, and 78% learned about their strengths and weaknesses. They (72%) also confirmed that the incomprehensible linguistic terminology was not common. Figure 2 summarizes the responses.
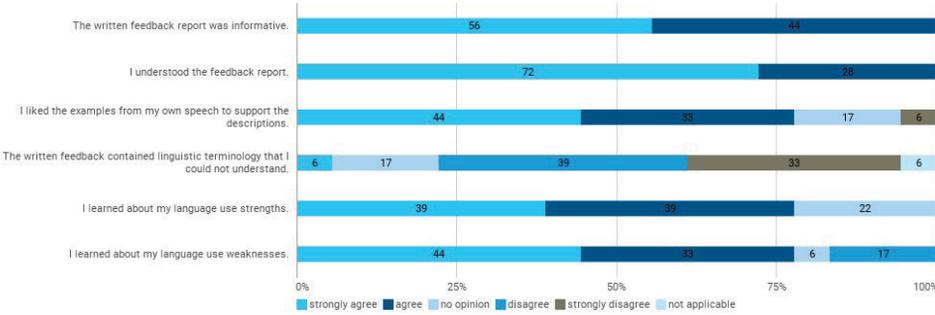

Figure 2. *Specificity and informative quality of the TOEPAS feedback report*

Most surveyed lecturers (98%) found it relevant, useful, and adequately detailed, and that it was a confirmation of their practice,

> Useful. There were some observations [about my speech] that I had never noticed before. (Respondent 1)
> It seemed relevant, constructive and included adequate detail. (Respondent 2)

When asked about recommendations for improvement, many (98%) had no suggestions or thought that the testing procedure and the feedback should remain the same, "No, it was fine, do not waste time on improving it" (Respondent 3). Some acknowledged the learning opportunities that the test offered and based on the feedback requested additional instruction.

Considering these learning opportunities, the initial expectation was that the oral feedback would be particularly useful for those with lower proficiency levels because they would be interested in more recommendations for improvement. However, lecturers who received top scores also showed interest in the oral feedback because it gave them an opportunity to discuss some issues they faced in the EMI classroom and obtain reassurance about their practices. In other words, the feedback helped them reduce their uncertainty about how well they perform (Ashford, Blatt and VandeWalle 2003). One lecturer stated that he appreciated "[t]he confirmation of [his] practice" (Respondent 4).

The effectiveness of the feedback is also supported by lecturers' surprise that the testing procedure was meaningful because their expectation was that it would be a mere formality. One lecturer, however, believed that lecturers should decide themselves whether to take the test or not in order to avoid possible perception of the procedure as "an accreditation exercise" (Respondent 5).

5 Conclusion

The results from the study suggest that language assessments for EMI certification can be transformed into instructional opportunities in order to support test-takers rather than to serve just as a tool for selection. Formative feedback, which normally is associated with students in instructional settings, can also become an integrated element of external assessment and offer benefits for all test-takers, including those who have less-privileged starting position, i.e. lower English proficiency.

In this study, the TOEPAS feedback balances facility and specificity in order to help test-takers understand the overall strengths and weaknesses and apply the recommendations beyond the test task. Feedback in instructional settings usually refers to the particular instructional task and provides guidance for revision. However, in test situations, despite references to the test performance, the purpose of the feedback is to facilitate improvement in real life contexts. Therefore, the lecturers appreciated that the feedback grounded their English language uses in the EMI domain and that they had the opportunity to discuss their challenges and uncertainties, as well as to seek recommendations in an oral feedback session. They tended not be threatened by the certification procedure when they were presented with the occasion to review the results and become more aware of the different aspects of their English language ability. In other words, to provide relevant feedback, test-takers' individual needs should be addressed, which could only be achieved with the test-developers' knowledge of the local context.

To conclude, a policy that forces the teaching staff to change the medium of instruction without any training support may not be defensible. A test based on such policy, therefore, might be considered unjust. However, using the test as an opportunity for learning, rather than exclusion, seems defensible, especially if benefits (learning opportunities) are available to everyone regardless of test results.

References

Airey, J. (2012). I don't teach language. The linguistic attitudes of physics lecturers in Sweden. *AILA Review,* 25: 64–79.

Ashford, S. J., Blatt, R. and VandeWalle, D. (2003). Reflections on the looking glass: A review of research on feedback-seeking behavior in organizations. *Journal of management,* 29(6): 773–799.

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

Ball, P. and 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 and D. Lasagabaster (eds.). *English-medium instruction at universities:Global challenges*, 44–64. Bristol, UK: Multilingual Matters.

Blattès, M. (2018). Policy development for English-medium instruction in French universities. *European Journal of Language Policy,* 10 (1): 13–37.

Davies, A. (2010). Test fairness: A response. *Language Testing,* 27 (2): 171–176.

Dempsey, J.V., Driscoll, M.P. and Swindell, L.K. (1993). Text-based feedback. In J.V. Dempsey and G.C.Sales (eds.). *Interactive instruction and feedback,* 21–54, Englewood Cliffs, NJ: Educational Technology Publications.

Deygers, B. (2019). Fairness and Social Justice in English Language Assessment. *Second Handbook of English Language Teaching*, 541–569.

Dimova, S. (2020). Language Assessment of EMI Content Teachers: What Norms. In M. Kuteeva, K.Kaufhold and N.Hynninen (eds.). *Language perceptions and practices in multilingual universities*. 351–378. London: Palgrave Macmillan.

Dimova, S. (2017). 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. and Kling, J. (2018). Assessing English medium instruction lecturer language proficiency across disciplines. *TESOL Quarterly,* 52(3): 634–656.

Dimova, S. and Kling, J. (2015). Lecturers' English proficiency and university language polices for quality assurance. In R.Wilkinson and M.L.Walsh (eds.). *Integrating content and language in higher education: From theory to practice. Selected papers from the 2013 ICLHE Conference,* 50–65. Frankfurt: Peter Lang.

Hattie, J. and Timperley, H. (2007). The Power of Feedback. *Review of Educational Research,* 77(1): 81–112.

Jensen, C. and Thøgersen, J. (2011). Danish University lecturers' attitudes towards English as the medium of instruction. *Ibérica,* 22:13-33.

Kling, J. (2013). *Teacher identity in English-medium instruction: Teacher cognitions from a Danish tertiary education context* (PhD dissertation). Copenhagen, Denmark: University of Copenhagen.

Kling, J., and Dimova, S. (2015). The Test of Oral English for Academic Staff (TOEPAS): Validation of standards and scoring procedures. In A.Knapp and K.Aguado (eds.).

Fremdsprachen in Studium und Lehre—Chancen und Herausforderungen für den Wissenserwerb, 247– 268. Frankfurt/Main: Peter Lang.

Kling, J. and Hjulmand, L.-L. (2008). PLATE–Project in language assessment for teaching in English. In R.Wilkinson and V.Zegars (eds.). *Realizing content and language integration in higher education,* 191–200. Maastricht: Universitaire Pers Maastricht.

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

Kulhavy, R.W. and Anderson, R.C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology*, (63): 505.

Mason, B.J. and Bruning, R. (2001). *Providing feedback in computer-based instruction What the research tells us. CLASS Research Report No. 9. Center for Instructional Innovation*. University of Nebraska-Lincoln.

McNamara, T. and Ryan, K. (2011). Fairness versus justice in language testing: the place of English literacy in the Australian citizenship test. *Language Assessment Quarterly,* (8): 161–178.

Messick, S. (1989). Validity. In: R.L.Linn (ed.). *Educational measurement* (3rd edn.)*,* 13–103. Washington, DC:American Council on Education/Macmillan.

Narciss, S. and Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H.Niegemann, R.Brünken and D.Leutner (eds.). *Instructional design for multimedia learning,* 181–195. Münster: Waxmann.

Phillipson, R. (2015). English as threat or opportunity in European higher education. In S. Dimova, A.K.Hultgren and C.Jensen (eds.). *English-medium instruction in higher education in* Europe, 19–42. Berlin: Mouton de Gruyter.

Phye, G. and Sanders, C.E. (1994). Advice and feedback: Elements of practice for problem solving. *Contemporary Educational Psychology*, (19): 286–301.

Pulcini, V. and Campagna,S. (2015). Controversy in Italian higher education. English-medium instruction in European higher education. In S.Dimova, A.K.Hultgren and C. Jensen (eds.). *English-medium instruction in higher education in* Europe, 65–88. Berlin: Mouton de Gruyter.

Rawls, J. (2001). *Justice as Fairness: A restatement*. Cambridge: Harvard University Press.

Santulli, F. (2015). English in Italian universities: The language policy of PoliMi from theory to practice. In S.Dimova, A.K.Hultgren and C. Jensen (eds.). *English-medium instruction in higher education in* Europe, 269–290. Berlin: Mouton de Gruyter.

Shute, V.J. (2008). Focus on Formative Feedback. *Review of Educational Research,* 78 (1): 153–189.

Surber, J.R. and Anderson, R.C. (1975). Delay-retention effect in natural classroom settings. *Journal of Educational Psychology*, (67): 170.

Underwood, J. (2008). Effective Feedback: Guidelines for Improving Performance. In G. Kanselaar, V. Jonker, P.A. Kirschner and F. J. Prins (eds.). *International Perspectives in the Learning Sciences:Creating a learning world. Proceedings of the Eighth International Conference for the Learning Sciences – ICLS 2008,* 415–422. Utrecht: International Society of the Learning Sciences.

Wilkinson, R. and Zegers, V. (2006). Assessing incipient linguistic competences: An institutional perspective. In R. Wilkinson and V. Zegers (eds.). *Bridging the assessment gap in English-medium higher education*, 61–76. Bochum: AKS-Verlag.

University of Copenhagen (2007). *Destination 2012: Strategy for the University of Copenhagen* [Online]. Available from: http://www.e-pages.dk/ku/236/ [Accessed: May 9th, 2020]