Curitiba – PR

Is there a minimum level of English proficiency to teach in a university EMI context?

Federal University of Paraná
University of Cambridge

Is there a minimum level of English proficiency to teach in a university EMI context?

Ron Martinez, Universidade Federal do Paraná (UFPR)

Sian Morgan, Cambridge Assessment English

ABSTRACT

As noted in recent surveys on EMI policy in higher education worldwide, there has been little to know empirical research into a recommended minimum level of proficiency to teach or learn in EMI contexts (Macaro et al., 2018; O'Dowd, 2018). In order to address this gap, it would be necessary to benchmark the proficiency of EMI lecturers and examine how those levels affect students' impressions of intelligibility. Thus, 22 Brazilian lecturers were recruited to take an EMI preparation course, and agreed to be recorded delivering a 20-minute lecture. 40 student raters were then recruited (including 13 non-Brazilian students) to use an adapted version of the Test of Oral English Proficiency in Academic Settings (TOEPAS). Results show that no lecturer below B2 was judged as ready to teach through English. At B2, 76% of all raters (n=40) agreed that this CEFR level was sufficient to teach through English, though often with some additional remediation recommended. (At C1/C2 there was 100% consensus that lecturers were EMI-ready.) The results can help inform future evidence-backed language policy decisions at Brazilian universities. Although the study has some limitations (e.g. only a 20-minute lecture sample, no longitudinal data), the B2 level – if confirmed as a meaningful threshold – is one that is achievable and realistic for the local scenario.

Key words: EMI, proficiency, intelligibility

INTRODUCTION

English as a Medium of Instruction (henceforth, EMI) in higher education settings in which English is not the usual language of instruction of the country is currently a growing global phenomenon (Dearden, 2014). Although the potential for EMI growth in Brazil is considerable (according to official data from the Brazilian Ministry of Education, there were 2,448 higher education institutions in existence as of October 2018[1]), it is still a relatively new trend in the country and, importantly, there are a number of barriers that may stifle its growth and expansion. In particular, there is a question concerning the adequacy of the level of proficiency of the majority of Brazilians to teach and learn through EMI. According to data from the now defunct Languages without Borders program which involved the testing of some 400,000 Brazilian students (and some faculty), less than 23 per cent of all candidates tested at a level equivalent to B2 or above[2]. By contrast, Klaassen and Bos (2010) report that over 45 per cent of academic staff at both Leiden University and Delft University in the Netherlands possess a level of proficiency at C1 or above, and O'Dowd (2018) cite this level as the required minimum for faculty to teach in English at over half of the 70 European universities surveyed.

However, there are at least two possible problems with asserting C1 as a minimum level of proficiency to teach in EMI contexts in Brazil. First, given the figures cited in the previous paragraph, it raises the question of how realistic it is for most Brazilians to reach such a standard. Second, and perhaps most important, there is no published empirical basis for such an assertion. As succinctly stated in a recent systematic review of EMI, "There is no higher education research, to our knowledge, that matches some kind of language proficiency test with actual practice in order to determine a minimum level to teach" (Macaro, Curle, Pun, An, & Dearden, 2018, p. 54).

It is nonetheless important to try to understand the extent to which 'proficiency matters' in an EMI context. After all, it can be argued that the expansion of internationalization in Brazilian higher education is at least in part dependent upon its ability to increase its engagement with the global academic community. This engagement, in turn, can be enhanced by offering courses through the language that is broadly considered an important academic lingua franca, i.e. English. It may or may not

---

[1] Source: http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/dados-do-censo-da-educacao-superior-as-universidades-brasileiras-representam-8-da-rede-mas-concentram-53-das-matriculas/21206

[2] Source: http://www.sri.cefetmg.br/2017/07/20/cefet-mg-obteve-bons-resultados-no-toefl-itp2016/ (accessed on June 11, 2019).

be the case that C1 – as is often cited – can be considered a recommended minimum to lecture through English in higher education, and if so then such a threshold greatly restricts the number of academic staff that can teach in EMI settings. If, however, lower levels of proficiency (e.g. B2) can be empirically shown to be acceptable by students, such a level may make the prospect of EMI expansion in Brazil more realistic. In any case, what is important is to go beyond 'guesswork'. Indeed, conjecture in itself may be harmful: language policy makers may set a bar that is unnecessarily too high (or even too low), and lecturers will not have a point of reference for which to strive (or reassure them).

To that end, the present study aimed to answer the following research question: *Is there a threshold of English proficiency, below which teaching in an EMI context becomes too burdensome for the EMI lecturer and/or students?*

METHOD

This study was conducted in the context of a Brazilian federal public university in the south of Brazil. Public universities are responsible for the largest share of research production in Brazil, holding the greatest number of post-graduate programs in the country. The growth of EMI in the public higher education sector has occurred chiefly in post-graduate education, and therefore the participants recruited for the present study were chosen especially with a view to representativeness of EMI as it currently stands in public tertiary education in Brazil.

Participants

There were two participant groups:

1. Lecturers – these were 22 Brazilian professors from various academic disciplines who opted into the main study, which took place at the UFPR campus in Curitiba, Brazil.
2. Students – there were 40 total students recruited for the study, broken down as follows: Brazilian (mostly post-graduate) (n=27), non-native English speaking non-Brazilian (n=9) from Germany, Colombia, Argentina, France, Mexico and Czech Republic, and native-speakers of English (n=4) from the United States.

The lecturers participated in an EMI faculty preparation course that was delivered by the Principal Investigator, assisted by a doctoral advisee whose thesis is related to teacher development in EMI. All lecturer participants opted into the study as a condition of participating on the course, and agreed to being recorded for the purposes of this research. In addition, all lecturers were required to take the *LinguaSkill* test (Cambridge Assessment English).

The Brazilian students were recruited from within a pool of students who had taken a post-graduate-level writing course delivered in English with the Principal Investigator, and only those who had declared their English to be "upper intermediate" or "advanced" were invited to participate. The non-Brazilian students were recruited on the first day of a special orientation that the office of international programs provides to new foreign students at the university. (In the case of the non-Brazilian participants, all were undergraduate students.). As their lecturer participant counterparts, all student participants were also required to take the *LinguaSkill* test.

The four native speakers were new arrivals on a Fulbright fellowship program (English Teaching Assistants). The native speaker were not asked to take the *LinguaSkill* test.

Materials

To benchmark proficiency, the *LinguaSkill* test was chosen. The *LinguaSkill* test is a computer-delivered and adaptive test. *LinguaSkill* was selected for this study since due to its ability to be delivered online, and for its reporting characteristics, which included CEFR-aligned scores, both overall and for all four assessed skills (reading, writing, listening and speaking).

In order to answer the research question, it would also be important to use a validated instrument to rate the English of lecturers. The Test of Oral English Proficiency for Academic Staff (TOEPAS) has been used extensively in EMI settings, and is arguably the only one of its kind, i.e. specifically designed to assess the proficiency of EMI lecturers. The TOEPAS was originally devised to be used by two raters who would use rubrics to determine a score of 1 to 5 (5 being equivalent to "native speaker") based on language-related criteria (pronunciation, grammar, vocabulary) and also teaching style (e.g. asking and answering student questions). The instrument is designed to have one agreed-upon rating that two raters provide (in the case of disagreement, a third rater can be included). However, for this study an adapted form of TOEPAS was used since the participants conducting the rating in this case would be non-language-specialist students, so some terms were made more 'user-friendly' for the participants (e.g. changing "lexis" to "vocabulary"). In the context of the original use of TOEPAS, a lecturer that receives a score of 4-5 is considered ready to teach EMI without further intervention, a score of 3 can be interpreted to mean acceptable to teach EMI, but with areas that can be improved, and scores 1-2 indicate the lecturer is not yet ready to teach EMI.

In addition, questionnaires were constructed for the lecturers and the students to measure their perceptions of their own proficiency, their language backgrounds, and learning/teaching backgrounds.

Procedure: Lecturers

The course the lecturers participated in was a one-week (5 days) intensive course with a total of 35 contact hours. One important role the course itself had in the design of the study was to decrease 'noise' in the eventual data produced that could result from broad heterogeneity in teaching dynamics. The EMI preparation course spends four days focusing on the importance of active methodologies (i.e. actively involving students, increasing teacher-student interaction), and so by the end of the course it is expected that all lecturers are delivering a style of class that can be characterized in a more uniform light.

The most relevant portion of the course (for the purposes of this study) occurred on the fourth day, in which lecturers began observed teaching. This teaching cannot be said to be reflective of an actual teaching scenario, for a number of reasons. First, the lecturers were only allotted 20 minutes each to deliver their 'classes'. Second, the lecturers were not teaching their usual students, but instead fellow course participants who, in most cases, were unfamiliar with the content that was being presented. Finally, and obviously, the mere fact of being observed cannot be said to be able to capture what a lecturer would typically do under unobserved circumstances.

Nevertheless, notwithstanding the aforementioned limitations, the design also offered a few advantages. Although the teaching environment itself in which this research was carried out cannot be said to ecologically valid, the very fact that all teaching occurred under such controlled circumstances also meant that the number of potential extraneous variables was easier to control. Thus, for example, elements such as time, classroom arrangement, number of students, and even lighting and acoustics were uniform throughout the research. Moreover, as already mentioned, the very fact that all lecturers had just undergone the same teaching methodology course also positively contributed to reducing heterogeneity and the influence of variables not being directly controlled for in the study.

All lecturers were video-recorded using high definition video with a separate lapel-mic audio feed, thus eliminating variability in audio quality that could stem from movement around the classroom. The position of the camera was always the same, moving only to follow the lecturer around if she or he walked away from the frame (for example, to monitor students in groups). Immediately following each recording, lecturers were requested to exit the room for a one-on-one interview with the Principal Investigator, which was also recorded.

All videos were edited using professional grade software, and interviews were transcribed.

Procedure: Students

PARTNER

BRITISH
COUNCIL   CONFAP
Conselho Nacional das Fundações
Estaduais de Amparo à Pesquisa

UK-Brazil
English**Collaboration**Call

A specially-designed website was created in order to deliver the video content to the student raters. Students were invited to attend a special rating session in a computer lab that was pre-prepared specifically for the rating exercise. Upon arrival, students were seated in pairs that had been strategically pre-determined with a view to optimizing the coverage of non-Brazilian students, especially (i.e. ensuring that these students viewed and rated as broad a range of CEFR bands as possible). Once all students had been seated, they were asked to begin the exercise by accessing the specially-designed website. This website contained three stages: 1) an introduction to the research, 2) a calibration exercise, and 3) the videos to be rated. After stage 1, students were given time to ask questions, which were promptly answered.

At stage 2, all pairs had the opportunity to view three different videos of three lecturers that had been identified by the PI as corresponding to different CEFR levels. After each viewing, students were required to first arrive at their own judgement and come to an agreement at the end. To do this, and to begin practicing using the instrument (TOEPAS), each student was given several copies of two documents: the TOEPAS (with space to write a score on top), and a sheet designed to keep notes. Students were told that both sheets would be collected at the end of the day. The student-raters' agreed-upon scores (1-5) were entered into a Google Form, and this form was monitored in real time by the PI in order to check that students were not rating lecturers too high or too low. (There were no cases identified in which students clearly rated a particular lecturer as too high or low.)

Finally, satisfied that students were able to use the TOEPAS instrument and were capable of rating as intended in the research design, all pairs rated their pre-determined video samples. All raters were required to view three lecturers (each 'class' lasting 20 minutes), stopping after each viewing to agree upon a score for that lecturer. During this time, the PI monitored activity, making sure all equipment was working properly, and so on.

When student-raters completed all three assessments, they were invited to be interviewed about their impressions of the lecturers, and of the rating exercise itself. These interviews were recorded and transcribed.

Analysis

All numerical data were entered into SPSS, which included the student LinguaSkill scores, the lecturer LinguaSkill scores, questionnaire data, and TOEPAS ratings.

RESULTS

The CEFR profile of the lecturers and students is provided in Table 1.

Table 1. Comparison between Lecturer and Student CEFR (as tested by LinguaSkill)

CEFR	Lecturers	Students
A1	4.55%	0
A2	4.55%	2.78%
B1	13.54%	5.56%
B2	63.54%	63.89%
C1+	13.54%	27.78%

Overall scores on *LinguaSkill* revealed that students' proficiency (M=**172.72**, SD=11.15) was slightly higher than that of the lecturers (M=**165.32**, SD=19.57), though this difference was not statistically significant (Kruskal-Wallis). Nonetheless, Table 1 shows that relatively few students were tested at the lowest CEFR ranges. Lecturers' scores on the LinguaSkill test had no significant interaction with age (Range: 29-60yrs, M=44.21). Lecturers had also been asked to rate their own proficiency in the background questionnaire. There was overall low agreement between self-reported and actual CEFR (κ = .161), but never by more than one band (e.g. B1 v. B2).

Within the individual skills measured by the test, listening had the strongest correlation with student ratings (r_s = .660, p <.01), though all skills were significant.

Most lecturers received three separate ratings (i.e. three ratings by different student pairs). Moreover, inter-class correlation revealed a high degree of agreement among all raters (**ICC** .804 ($F(15,32)$=5.096, *p*<.001), with no significant difference related to nationality (i.e. Brazilian or Non-Brazilian) or being a native speaker of English. A summary of the ratings of all lecturers (n=22) is provided in Table 2.

Table 2. Summary of ratings of lecturers as given by student raters

CEFR	Number of lecturers	Number of paired ratings	"NOT READY FOR EMI" Number of pairs that gave a score of **1**	"NOT READY FOR EMI" Number of pairs that gave a score of **2**	"READY FOR EMI" Number of pairs that gave a score of **3**	"READY FOR EMI" Number of pairs that gave a score of **4**	"READY FOR EMI" Number of pairs that gave a score of **5**	Total percent score of ≥ 3
A1-A2	2	7	3	3	1	0	0	14%

B1	3	9	3	5	1	0	0	1%
B2	14	0	1	9	17	11	2	6%
C1+	3	7	0	0	1	4	2	00%
TOTAL	22	63	7	17	20	15	4	

DISCUSSION

The data in Table 2 can be interpreted in a number of ways. If one were to point to a CEFR level that appears to be a 'safe bet' for EMI, then clearly C1 stands out. However, considering the research question ("Is there a level of proficiency below which teaching in EMI becomes to burdensome?"), then the answer would seem to be B2. Naturally, there is a caveat of sample size (14 lecturers tested at B2 versus only 5 below and only 3 above), but at least at the B2 level there would seem to be sufficient data to indicate that a majority of student raters – 76% -- agree that lecturers tested at this level possess a level of proficiency that is acceptable to teach in an EMI context.

There was not, however, a consensus among those lecturers (contrary to those tested at "C1+"). Thus, it is worth looking at those lecturers who received lower ratings to explore what set them apart. It may also be interesting to examine the lecturers that were tested at B2 but received extraordinarily high ratings. These lecturers are profiled in Table 3.

Table 3. Lecturers at B2 with extraordinarily low or high ratings by students

NAME	CEFR	LIST.	READ.	SPK.	WRIT.	TOTAL	R1	R2	R3	Mean Rating
LENNY	B2	161	180	158	153	**164**	2	2	2	2.00
BETTY	B2	169	168	X	159	**165**	2	3	2	2.33
MARY	B2	170	180	180	169	**175**	1	2	4	2.33
SALLY	B2	170	178	156	171	**169**	2	3	X	2.50
SUNNY	B2	159	164	168	162	**163**	3	2	X	2.50
ELLY	B2	180	179	178	164	**177**	3	5	5*	4.33
ALLY	B2	180	178	174	165	**176**	4	4	X	4.00

Interestingly, the skill that most stands out as different between the two types of B2 lecturers profiled in Table 3 (i.e. extraordinarily low or high ratings) is listening. It should be noted that the score of 180 is a maximum score – none of the lecturers who were rated low (i.e. mean rating below 3) by students reached that maximum score. This finding is consistent with Wagner (2016), who found that TOEFL iBT listening scores correlated most robustly with student ratings of international teaching assistants in a North American higher education context.

However, to better understand what could be driving the ratings down of the handful of B2 professors shown in Table 3, and what could, conversely, be associated with higher ratings of B2 lecturers, the qualitative data (i.e. interviews) were also analyzed.

In the analysis of the B2 lecturers that received especially lower ratings, the comments that were most salient were related to dysfluencies (e.g. pauses), and also the presence of Portuguese, either by way of accentedness, or lexis/grammar, or both. By contrast, the B2 lecturers that received especially higher ratings were often characterized as seeming "comfortable" with English, which meant no or few language-related pauses, and also no use of Portuguese.

CONCLUSION

To date there had been no empirical research to indicate a proficiency threshold below which teaching through English becomes too burdensome for the teacher and/or student in EMI settings. The data in our study indicate that B2 is a meaningful threshold. At this level, the majority of student raters (76%) agreed that the lecturer tested at B2 was acceptable to teach an EMI class. If taken at face value, the fact that most lecturers tested at B2 were considered acceptable for EMI is an encouraging finding, since this CEFR level is one that could be considered more realistic for the Brazilian context.

However, the research also points to a number of questions. First, given the sample size, one should exercise caution in asserting that B2 definitely represents a meaningful threshold. The current study can be said to provide good evidence for this assertion, but by no means conclusive. Moreover, the most that could be claimed with any degree of confidence is that most lecturers at B2 were found to be acceptable for a Brazilian context – more specifically, a post-graduate public university EMI context.

Moreover, as shown in Table 3, 5 out of the 14 lecturers tested at B2 were judged as lower than the majority of the cohort at that level. Yet a number of possible confounding variables may be at play here. Crucially, it is important to remember that the lecturers were being rated by students that were not from the lecturers'

disciplines, and the lecturers had not only never delivered a class through English before, but for many of them, the occasion on which they were recorded also represented the first time they had attempted to teach in an interactive lecture delivery style. As such – and taken together with the fact that they were being recorded and had unusual time pressure (i.e. only 20 minutes) – a serious question can be raised concerning the validity of their data. Hence, one important future direction for further study in this line of inquiry would be to use more longitudinal data in more ecologically valid settings. This could occur, for example, by repeating the research exercise described here but with lecturers who are teaching their actual students, in intact classes. A measure could be applied at the beginning of their teaching (roughly tantamount to the current study), and then apply a repeated measure at a later date, closer to the end of the term, in order to measure any possible change in rating. Such a design would help to validate the findings reported in this study.

REFERENCES

Ball, P., & Lindsay, D. (2013). Language demands and support for English-medium instruction in tertiary education. Learning from a specific context. *English-medium instruction at universities: Global challenges*, 44-64.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600-1610.

Choi, I. (2017). Empirical profiles of academic oral English proficiency from an international teaching assistant screening test. *Language Testing*, *34*(1), 49-82.

Dearden, J. (2014). English as a medium of instruction–a growing global phenomenon. *British Council. Online: http://www. britishcouncil. org/education/ihe/knowledge-centre/english-language-higher-education/report-english-medium-instruction (accessed 9 October 2015)*.

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.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language learning*, *34*(1), 65-87.

Grift, Y. K., Meijer, A. A., & van der Salm, F. H. (2012). The impact of the language of instruction: How economics students in the Netherlands evaluate an English-taught undergraduate programme. *Discussion Paper Series/Tjalling C. Koopmans Research Institute*, *12*(2).

Halleck, G. B., & Moder, C. L. (1995). Testing language and teaching skills of international teaching assistants: The limits of compensator strategies. *TESOL Quarterly*, *29*(4), 733-758.

Hendriks, B., Van Meurs, F., & Hogervorst, N. (2016). Effects of degree of accentedness in lecturers' Dutch-English pronunciation on Dutch students' attitudes and perceptions of comprehensibility. *Dutch Journal of Applied Linguistics*, *5*(1), 1-17.

Johncock, P. (1991). International Teaching Assistants Tests and Testing Policies at US Universities. *College and University*, *66*(3), 129-37.

Kang, O., Rubin, D., & Lindemann, S. (2015). Mitigating US undergraduates' attitudes toward international teaching assistants. *Tesol Quarterly*, *49*(4), 681-706.

Kang, O., Thomson, R., & Moran, M. (2018). The Effects of International Accents and Shared First Language on Listening Comprehension Tests. *TESOL Quarterly*.

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*(2010), 61-75.

Kling, J., & Dimova, S. (2015). The Test of Oral English for Academic Staff (TOEPAS): Validation of standards and scoring procedures. In *Fremdsprachen in Studium Und Lehre-Chancen Und Herausforderungen Für Den Wissenserwerb*(pp. 247-268). Peter Lang.

Kling, J., & Stæhr, L. S. (2012). The development of the test of oral English proficiency for academic staff (TOEPAS). *University of Copenhagen*.

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, *51*(1), 36-76.

Margić, B. D., & Vodopija-Krstanović, I. (2018). Language development for English-medium instruction: Teachers' perceptions, reflections and learning. *Journal of English for Academic Purposes*.

Martinez, R. (2016). English as a Medium of Instruction (EMI) in Brazilian higher education: challenges and opportunities. In K. Finardi (Ed.) *English in Brazil: Views Policies and Programs*, (pp. 191-228). Londrina: Eduel

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in second language acquisition*, *28*(1), 111-131.

O'Dowd, R. (2018). The training and accreditation of teachers for English medium instruction: an overview of practice in European universities. *International Journal of Bilingual Education and Bilingualism*, *21*(5), 553-563.

Schmidgall, J. E. (2013). *Modeling speaker proficiency, comprehensibility, and perceived competence in a language use domain* (Doctoral dissertation, UCLA).

Wagner, E. (2016). A Study of the Use of the TOEFL iBT® Test Speaking and Listening Scores for International Teaching Assistant Screening. *ETS Research Report Series*, *2016*(1), 1-48.

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