

Pragmatic outcomes in the English-medium instruction context

The influence of intensity of instruction

Ana Herraiz-Martinez and Eva Alcón-Soler

Universitat Jaume I

This study investigates pragmatic development in the English-medium instruction (EMI) setting of the Valencian Community in Spain. More specifically, the study examines whether the intensity of EMI influences functional adequacy (FA) in second language (L2) writing. Participants were 102 EMI learners, each of whom wrote three motivation letters over one academic year in English. The rating scales designed by Kuiken and Vedder (2017) were used to examine the FA of the written texts in terms of cohesion, coherence, task requirements, content, and comprehensibility. Quantitative results revealed significant differences among the EMI groups under analysis, suggesting that the intensity of instruction may exert an influence on FA in L2 writing. Results from this study show the importance of intensity of exposure to EMI for L2 writing.

Keywords: English-medium instruction, EMI, functional adequacy, pragmatics, intensity of exposure

1. Introduction

Research on classroom teaching and learning of pragmatics has increased greatly in the last three decades in traditional second language (L2) instructional contexts (see Taguchi, 2015, for an overview of instructional studies in pragmatics). However, new English language environments have emerged, and there is a need to conduct research across new contexts (Bardovi-Harlig, 2013).

One learning environment that has motivated research in the last decade is the English-medium Instruction (EMI) classroom. The influence of EMI, not only in language learning in general but also in pragmatics, has been emphasized by a number of scholars. For instance, Wong (2010, p.126) states that EMI contexts are ideal teaching environments since “students will become more comfort-

able in an English-rich communication forum, helping them to perceive English less as a dry 'theoretical' subject and more as a vibrant, living means of dialogue". Similarly, Taguchi (2012) and Ohta (2001) explain that, although EMI contexts typically do not have clearly defined linguistic outcomes, incidental learning of linguistic forms may occur in these contexts due to the large amount of target language exposure students receive. With this in mind, it might be inferred that EMI contexts have one main advantage over foreign language (FL) contexts: their role as facilitators of language learning and communication. Despite the increase in studies in EMI settings, research to date has reported the advantages of EMI on students' linguistic abilities only (Ament & Pérez-Vidal, 2015; Loranc-Paszylk, 2007), and students' functional abilities, particularly pragmatics, has not received much attention as learning outcomes of EMI (e.g., Taguchi, 2012; Taguchi, Naganuma & Budding, 2015). Given the scarcity of longitudinal studies on pragmatics learning in EMI contexts, the present study explores whether the intensity of EMI (50%, 75% or 100% of instruction in English) influences functional adequacy (FA) in L2 writing.

The paper is structured as follows. First, the background section provides an overview of research on pragmatics in EMI contexts and intensity of exposure, and synthesizes the defining features of FA, which is the pragmatic target of this study. Next, the method section provides information about research participants, data collection instruments, and data analysis methods used in the present study. After that, results are presented. This section is followed by a discussion of results, and the paper finishes highlighting the main conclusions, limitations and pedagogical implications.

2. Background research

2.1 Written pragmatic production in EMI settings

Research on pragmatics learning in the EMI setting is still in its initial stages (e.g., Taguchi, 2012; Taguchi et al., 2015). As far as we are aware, only two recent studies have addressed written pragmatic production in EMI contexts: Ament and Pérez-Vidal (2015) and Salaberri and Sánchez-Pérez (2015). The study conducted by Ament and Pérez-Vidal (2015) constitutes an attempt to assess L2 learners' pragmatic errors through a written composition. Sixteen students participated in this longitudinal pre-test-post-test experimental study over one academic year in a university in Spain. They were split into a full immersion group (entire degree in EMI) and a semi-immersion group (half of their degree through EMI). The authors examined learners' overall linguistic gains (listening

and lexico-grammatical abilities). They also analyzed the written task according to coordination, accuracy and fluency measures. Accuracy measures involved calculating the number of pragmatic errors per clause, which included errors on referents and discourse connectors, as well as incorrect use of idioms, expressions and formulaic language. Results revealed a slight improvement for accuracy measures in the full immersion group, as fewer errors were detected in the post-test task. As for the semi-immersion group, results indicated no changes in regards to pragmatic errors per clause since accuracy measures remained the same over the academic year.

Salaberri and Sánchez-Pérez (2015) analyzed L2 written production of 67 learners enrolled in English courses in the Agricultural Engineering degree in Spain. Participants were required to produce, in written English, a specific assignment based on an experiment task conducted in a laboratory, as well as a final lab report. The rating scale used to assess learners' written production addressed the degree of task fulfillment, organization of content, grammar and vocabulary. Task fulfillment analyzed the degree of appropriateness and register (directly relevant to pragmatics), while text organization assessed the extent to which the communicative purpose of the reports was achieved (use of discourse markers and overall structure of discourse). Results showed differences in learners' performance across areas assessed: Acceptable performance was found regarding the command of grammar and vocabulary, but weaknesses were found in discourse-related aspects of performance (e.g., overall structure). Accordingly, the researchers argue for "the need to incorporate in EMI classrooms new methodologies that help students integrate the global features of the writing ability within their own course contents in a second language [...]" (Salaberri & Sánchez-Pérez, 2015, p. 55).

2.2 The role of intensity of exposure in English-medium instruction

Scholars generally agree that there is a close relationship between language learning and intensity of exposure to the target language (TL). For instance, Sheela & Ravikumar (2016, p. 772) claim that "students with more exposure to the TL are expected to acquire greater familiarity with the target language". This implies that while L2 learners are in the process of learning the TL, they can acquire more aspects of the language if they have more contact with it, albeit in verbal or written form, or in formal or informal ways of communication. In the case of EMI, the relationship seems to be stronger given that the exposure to English is the main trait of this instructional approach. Indeed, since EMI provides a suitable environment to enhance learners' exposure to English, one may hypothesize that learning outcomes differ corresponding to the amount of exposure they receive in this formal setting. For example, a program offering 270 hours of EMI would not provide

students with the same opportunities as one of 600 hours of EMI; that is, more exposure to TL could lead to more opportunities to express ideas in English and to acquire specific knowledge in English.

Although this hypothesis is plausible, empirical research on how the intensity of English instruction influences pragmatics learning is still in its infancy. To the best of our knowledge, only one study, Zhang and Yang (2012), has addressed this issue. These authors analyzed the impact of the number of hours of English language instruction on 128 Chinese English as a Second Language (ESL) learners' sociopragmatic awareness. The authors found that learners who had studied five main subjects through English significantly improved their pragmatic competence than those who studied fewer subjects. In particular, there was a significant effect of intensity of instruction in the speech act of request. The researchers attributed these findings to the naturalistic communicative situations existing in EMI classes: Learners were encouraged to use English for authentic communicative purposes through activities such as games and chants. Moreover, as the researchers claim, since teachers were required to speak only in English, the learning of pragmatic norms is likely to take place when learners interact with their teachers on meaningful contents. Hence, findings from Zhang and Yang's study present a counter argument to Schmidt's (1993) claim that mere exposure to the TL does not automatically result in pragmatics learning.

Other studies have explored the role of increased exposure to the TL in pragmatic development (e.g., Sánchez-Hernández, 2017) and in general language development (not specific to pragmatic development) (Ament & Pérez-Vidal, 2015; Housen, 2012; Lightbown & Spada, 1991; Loranc-Paszylk, 2007; Serrano & Muñoz, 2007). These studies have explored how intensive exposure to TL affects language learning in formal settings and they provide relevant findings for the purpose of the present study. As a matter of fact, they contribute to our understanding of how the amount of EMI exposure influences linguistic outcomes in structured and purposeful classroom learning contexts.

Loranc-Paszylk (2007) examined undergraduate students' development of linguistic abilities after an EMI treatment in Poland. Using a reading task, the researcher compared the performance of two EMI programs, i.e., English Philology and International Relations. The latter program offered 60% less English exposure than the former program. Results revealed that the performance of both groups was comparable. These findings contradict with results in Housen's (2012) study, which found greater language development corresponding to the greater amount of input available in intensive instructional settings.

Acknowledging the lack of studies examining the effects of EMI on linguistic performance, Ament and Pérez-Vidal (2015) have recently addressed this research gap by measuring linguistic gains of EMI students using four tasks, i.e., an oral

comprehension task, a written composition, a cloze task, and a grammar task. This study adopted a longitudinal pre-test-post-test experimental design over one academic year. Two groups of participants were involved: (1) the full immersion group who had 100% exposure to EMI (1500 hours by the end of the academic year), and (2) the semi-immersion group who had 18–41% exposure to EMI (275–625 hours). Findings suggested that, although both groups improved on their linguistic performance, the full immersion group showed larger gains.

Further evidence on the influence of exposure on language gains comes from studies that revealed significant differences between different types of immersion programs. For instance, Burmeister and Daniel (2002) examined the effect of partial immersion programs by collecting students' oral and written production data. The researchers analyzed the influence of TL input on the length of students' production in terms of the number of clauses, the absolute frequency of cohesive elements used as well as their density. In the partial immersion program, classes were taught about 30% in English (15% more English exposure compared with a regular curriculum in the same university). Results showed that the students in the immersion program scored higher than those in the regular curriculum with respect to the number of clauses and the frequency of cohesive devices in their production. These findings suggest that more exposure via EMI is more beneficial for language learning.

In summary, it seems that L2 learners in EMI programs may experience language learning to different extents with regard to the intensity of exposure to the TL. Although several studies have reported findings on the role of exposure in immersion by comparing different programs, there is still an important research gap to fill, since, with the exception of Zhang and Yang's (2012) study, the role of intensity of EMI on pragmatics learning has not been investigated. The present study intends to fill this gap by examining how different levels of intensity of EMI affect L2 learners' pragmatic development over one academic year.

2.3 Functional adequacy and pragmatics

The central question in SLA research is what makes L2 learners proficient users of the TL. Different terminologies have been used to describe learners' proficiency, including communicative adequacy (Kuiken et al., 2010; Pallotti, 2009), communicative competence (Bachman & Palmer, 1996), intercultural competence (Usó-Juan & Martínez-Flor, 2006), communicative effectiveness (Sato, 2012), and communicative functionality (Fragai, 2001, 2003). Functional adequacy is another term used to describe L2 proficiency.

Functional adequacy has been interpreted in terms of coherence and cohesion of text (Knoch, 2009), socio-pragmatic appropriateness (McNamara & Roever,

2007), and successful transfer of information (Upshur & Turner, 1995). However, Kuiken and Vedder (2017, p. 323) use functional adequacy to mean “successful task fulfillment”. As such, functional adequacy is viewed as a task-related, interpersonal construct, involving two participants collaboratively constructing meaning while completing a task. In addition, functional adequacy is a context-dependent construct because tasks are performed in specific institutional settings where language learning takes place (e.g., EMI or CLIL). Given these specifications, Kuiken and Vedder’s (2017, p. 323) original definition of functional ability can be elaborated as follows:

Functional adequacy is viewed as a task-related, interpersonal construct, involving two participants (Participant A and Participant B) collaboratively constructing meaning. Functional adequacy is determined in terms of participants’ successful task fulfillment, in which Participant A conveys message and Participant B understands the message while completing a task.

We argue that functional adequacy corresponds to pragmatic competence because it involves ability to convey meaning effectively to complete a communicative task in a specific context.

The purpose of the present study is to assess development of functional adequacy in L2 writing in an EMI setting in Spain. Since functional adequacy is the main focus of this study, empirical studies assessing this construct are reviewed in the next section.

2.3.1 *Studies assessing the functional dimension of the language*

The last two decades have seen a flourishing body of research investigating L2 learners’ task-based performance from the perspective of communicative/functional adequacy (e.g., Kuiken et al., 2010; Martin-Laguna, 2018; Revesz, Ekiert & Torgersen, 2016). Two major research projects have characterized this trend: (1) the *What is Speaking Proficiency* (WISP) and (2) the *Communicative Adequacy and Linguistic Complexity* (CALC). Although both examined communicative/functional adequacy of L2 production, the focus of investigation differs: The former focuses on spoken production, while the latter focuses on written production.

Within the WISP project, two recent studies are worth mentioning. De Jong et al. (2012a) investigated the extent to which L2 knowledge and processing skills could explain L2 speaking proficiency (specifically conceptualized as functional adequacy). Results showed that vocabulary knowledge and intonation were the strongest indicators of speaking proficiency, as they predicted 75% of speaking proficiency. On the other hand, Revesz et al. (2016) recently examined the influence of task type on functional adequacy and linguistic outcomes. Results indicated that task type did not have any effect on functional adequacy or measures of

complexity, accuracy, and fluency (CAF). Results also showed that the dominant predictor of communicative adequacy was the frequency of filled pauses (fluency of speech).

Within the CALC project, Kuiken et al.'s (2010) study examined learners' functional adequacy on written production. This study involved 94 L2 learners of Dutch, Italian and Spanish, whose proficiency levels ranged from A2 to B1 according to the Common European Framework of Reference (CEFR). All participants completed two open-ended decision-making tasks, which were rated by native speakers (NSs) and L2 teachers of the corresponding TL. Results showed that functional adequacy and linguistic complexity of written samples developed at an equal pace. In addition, these variables correlated significantly, especially for advanced-level learners. The most notable contribution of the CALC project relates to the development of a valid and reliable rating scale assessing functional adequacy. The holistic rating scale comprised seven different levels and was used to rate learners' performances based on general descriptors of accuracy, syntactic complexity, and lexical complexity.

The studies discussed above demonstrate the importance of considering communicative adequacy as a crucial dimension of L2 proficiency (Fulcher, 1987; Knoch, 2011), as well as the need to assess functional adequacy of learners' production separately from CAF measures. Inspired by Grice's (1975) maxims of quality, quantity, relevance, and manner, Kuiken and Vedder (2017) proposed a new rating scale for assessing functional adequacy of L2 written production. Based on the general proficiency descriptors provided by the CEFR (Council of Europe, 2001) and the rating scale suggested by De Jong et al. (2012a, 2012b), Kuiken & Vedder (2017) proposed a scale assessing four main dimensions: (i) content, (ii) task requirements, (iii) comprehensibility, and (iv) coherence and cohesion. The authors claimed that these four dimensions together determine learners' functional adequacy. Reliability and validity of the scale were fully tested in the CALC project.

Kuiken and Vedder (2017) argue that, in order to investigate the generalizability of the findings, the scale should be tested with participants of different educational backgrounds, task types, and target languages other than Dutch and Italian. However, to our knowledge, only one study examined functional adequacy of L2 English learners in an EMI setting. Herraiz-Martinez and Alcón-Soler (2018) conducted a longitudinal study tracing development of functional adequacy of English compositions written by EMI learners in tertiary education. The study found that learners' performance improved during the EMI experience over one academic year. However, this study did not compare different EMI groups in terms of their intensity of instruction. To further examine the applicability of functional adequacy in L2 production, this study adapted Kuiken and Vedder's (2017) rating

scale to assess functional adequacy of learners' written production in three groups of differing EMI intensity: 50%, 75% and 100% of EMI instruction.

3. Method

3.1 Participants

A total of 102 EMI students (44 males and 58 females) from two state-run higher education institutions in the Valencian Community participated in this study. Their ages ranged from 17 to 25 years old. Their English proficiency was judged to be upper intermediate based on the standardized Quick Oxford Placement test (UCLES, 2001). Participants were enrolled in three different bachelor EMI programs in the first of a four-year non-compulsory stage of tertiary education. The EMI programs differed in terms of the degree of exposure to English as the medium of instruction. More specifically, 39 students were enrolled in the English Studies degree (50% of the credit hours in English, i.e., 900 hours total), 32 in the Economics degree (75% of the credit hours in English, i.e., 1350 hours total) and 31 in the Biotechnology degree (100% of the credit hours in English, i.e., 1800 hours total).

3.2 Data collection instruments

In the present study, a motivation letter (a cover letter in which students explain why they are qualified candidates for the post) was used to collect data for three main reasons. First, this type of text is a requirement for the students wishing to participate in Erasmus programs or making an application to a specific company to carry out the internships. Second, motivation letters are persuasive texts that carry certain pragmatic force. These letters have direct effect on audience and thus reflect real-life language use (see Gomez-Laich & Taguchi, 2018, for the use of persuasive writing in examining L2 English learners' task-based pragmatic performance). Third, motivation letters were used in order to avoid participants' familiarity with other types of tasks such as opinion essays, which they practiced writing in their previous educational stages (secondary education).

Students from the three EMI groups were asked to write motivation letters to request an opportunity to conduct internships in a company related to their field of study. They repeated the same task three times during the academic year. In order to improve the authenticity of the task, we included an invitation letter from the president of each of the companies addressed in the letters, with a brief description of the company (i.e., number of employees and location of the headquarters) and a

number of available internship places. Accordingly, the president encouraged participants to write a motivation letter to the recruitment team who would examine the possibility of engaging their services. Students were given 10 minutes to go over all the information and details provided in the president's letter. Then, they were asked to write a letter within 25 minutes (130–150 words in length).

3.3 Data analysis

A total of 306 motivation letters were analyzed. In other words, each participant wrote three compositions in English over one academic year. Data were gathered in three waves (Time 1, Time 2 and Time 3) and there were three months of instruction between each data collection phase. The essays were typed up by the researchers exactly as they were written in the paper format by learners (leaving mistakes uncorrected) for further analysis.

A modified version of Kuiken and Vedder's (2017) rating scale was used to assess writing samples (See Appendix). The scale contained separate analytical rubrics to assess learners' functional adequacy (FA) through the holistic scoring of specified dimensions. The six-point scale proposed by Kuiken and Vedder was inspired by Grice's (1975) conversational maxims, focusing on quantity, relevance, manner, and quality of the written message. Accordingly, four dimensions were comprised in the rating scale: (i) content, (ii) task requirements, (iii) comprehensibility, and (iv) coherence and cohesion (see Kuiken & Vedder, 2017, for the reliability and validity of the scale). Although cohesion and coherence are two properties of texts that contribute to the overall text interpretation, in the present study, coherence and cohesion were treated as two separate dimensions in case some letters are cohesive but not coherent. For example, some paragraphs were comprehensible, but did not have any cohesive devices (see Example (1)). On the other hand, some sentences were connected with several discourse markers (i.e., cohesive devices) but were not coherent (see Example (2)).

- (1) I'm (name of participant). I'm 23 years old. I live in Petrer (Alicante). I have two sisters. Their names are X and Y. (#68/EMI2/PRE#)
- (2) **To conclude**, I want to say that I'm different teacher. I don't use books in my Spanish classes **because** I think that doesn't the good option for children to learn and be. (#13/EMI0/PRE#)

By separating cohesion and cohesiveness, the present study measured FA on five dimensions: (i) coherence, (ii) cohesion, (iii), comprehensibility, (iv) content, and (v) task requirements. These five dimensions collectively address learners' pragmatic competence because pragmatic competence involves ability to produce functionally adequate texts when performing a pragmatic-focused task (i.e.,

persuading someone). Each dimension was assessed on a five-point scale and thus was given a score of 0, 1, 2, 3, 4 or 5 (see Appendix). This scoring procedure was used at Time 1, Time 2, and Time 3 for a total of 306 motivation letters. A spreadsheet was created in order to perform the corresponding analyses and to explore learners' development in each of the dimensions (content, coherence, cohesion, comprehensibility and task requirements) over the academic year.

The motivation letters were rated by the researchers and two secondary-school English teachers in order to ensure consistency and objectivity in assessing learners' FA. These raters went through six training sessions to familiarize themselves with the rating scales and assessment procedures. They independently rated 60% of the letters (243 letters). The inter-rater agreement rate was 88% for coherence, 87% for cohesion, 91% for comprehensibility, 90% for task requirements, and 89% for content. Multivariate analysis of variance (MANOVA) was used to compare learners' performance across three EMI groups and time points. The independent variable was time with three levels (Times 1, 2 and 3), and the dependent variable was the score obtained for each of the five dimensions in the assessment of FA.

4. Results

The research question addressed the development of learners' FA across three EMI groups over time. Table 1 displays the descriptive statistics of average score on the coherence dimension of the motivation letters. T1, T2, and T3 refer to three time points (Time 1, Time 2, and Time 3). Groups 1, 2 and 3 refer to the three EMI groups (50%, 75% and 100% of EMI intensity). Figure 1 further illustrates the descriptive findings from Table 1.

Table 1. Descriptive statistics of coherence performance

	N	T1		T2		T3	
		M	SD	M	SD	M	SD
Group 1 (50%)	39	2.49	0.82	3.72	0.94	4.33	0.77
Group 2 (75%)	32	2.13	1.04	3.34	0.90	4.13	0.90
Group 3 (100%)	31	2.94	1.43	4.32	1.10	4.65	0.60

By observing Table 1 and Figure 1, we can see that the three groups improved their ability to write coherent texts over the academic year. MANOVA revealed significant group differences at T1 [$F(2,99)=4.241$; $p=.017$], T2 [$F(2,99)=7.931$; $p=.001$], and T3 [$F(2,99)=3.604$; $p=.031$]. Post-hoc Tukey and Games-Howell comparison revealed that Group 3 (100% EMI) was significantly better than

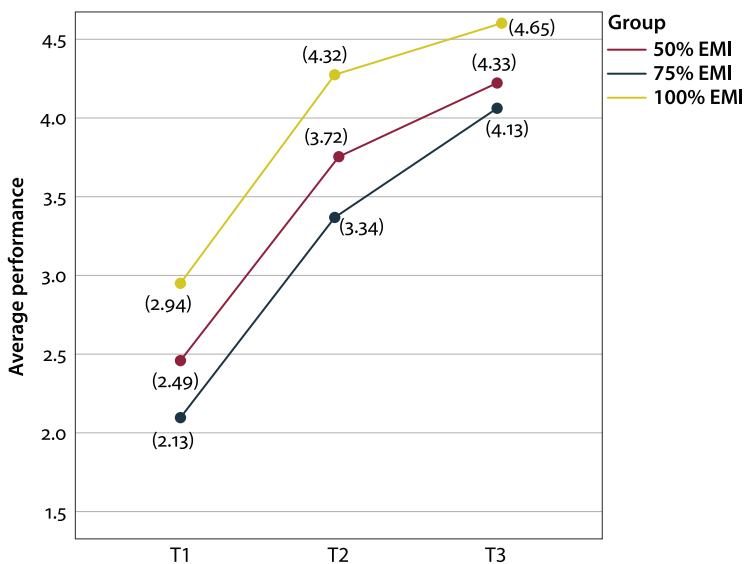


Figure 1. Coherence performance across groups

Group 1 (50% EMI) at T2 (mean difference = 0.60, $p = .032$) but not at T3 (mean difference = 0.31, $p = .149$). The tests also revealed that Group 3 (100% EMI) was significantly better than Group 2 (75% EMI) at T1 (mean difference = 0.81, $p = .012$), T2 (mean difference = 0.98, $p < .001$) and T3 (mean difference = 0.52, $p = .026$). These findings suggest that the performance of coherence was affected by intensity of instruction (i.e., 100% EMI).

Table 2 displays the descriptive statistics of average score on the cohesion dimension of the motivation letters. Figure 2 further illustrates the descriptive findings from Table 2.

Table 2. Descriptive statistics of cohesion performance

	T1			T2		T3	
	N	M	SD	M	SD	M	SD
Group 1 (50%)	39	2.26	0.93	3.46	1.09	4.23	0.93
Group 2 (75%)	32	1.66	1.18	2.91	0.92	4.03	0.89
Group 3 (100%)	31	2.55	1.54	4.00	1.00	4.29	0.78

From Table 2 and Figure 2, we can observe that the three groups improved their ability to produce cohesive texts over the academic year. The MANOVA test revealed significant group differences at T1 [$F(2, 99) = 4.391$; $p = .015$] and T2 [$F(2, 99) = 9.108$; $p < .001$]. However, such differences were not observed at T3 [$F(2, 99) = 0.767$; $p = .467$]. Post-hoc Tukey and Games-Howell comparisons

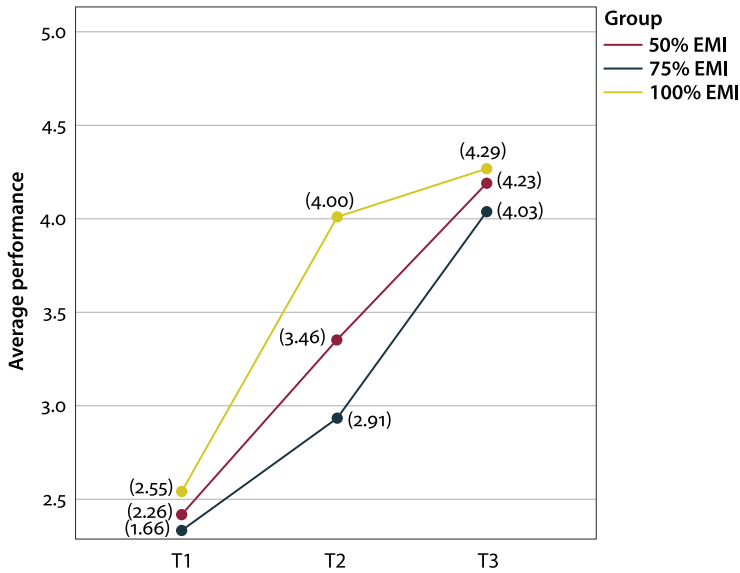


Figure 2. Cohesion performance across groups

revealed no significant difference between Group 3 (100% EMI) and Group 1 (50% EMI) at T1 (mean difference = 0.29, $p = .627$), T2 (mean difference = 0.54, $p = .076$) or T3 (mean difference = 0.06, $p = .957$). However, the tests revealed that Group 3 (100% EMI) was significantly better than Group 2 (75% EMI) at T1 (mean difference = 0.89, $p = .034$) and T2 (mean difference = 1.09, $p < .001$), but not at T3 (mean difference = 0.26, $p = .473$). These findings suggest that the effect of EMI intensity was not straightforward, and when there was effect, it was not maintained long term.

Dealing with the third sub-component of FA, Table 3 includes the descriptive statistics of average score on the comprehensibility dimension of the motivation letters. Figure 3 further illustrates the descriptive findings from Table 3.

Table 3. Descriptive statistics of comprehensibility performance

		T1			T2		T3	
	N	M	SD	M	SD	M	SD	
Group 1 (50%)	39	2.87	0.92	3.95	0.88	4.59	0.59	
Group 2 (75%)	32	2.50	0.98	3.50	0.84	4.09	0.92	
Group 3 (100%)	31	3.13	1.23	4.32	0.90	4.55	0.62	

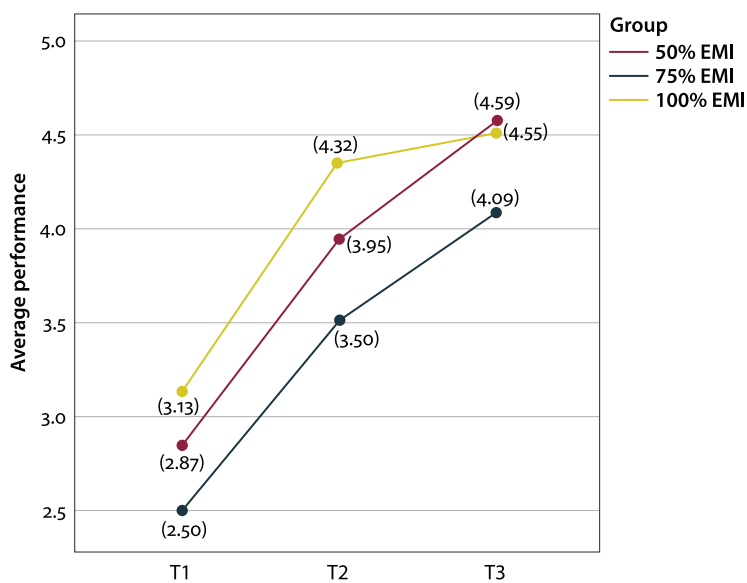


Figure 3. Comprehensibility performance across groups

As shown in Table 3 and Figure 3, we can observe that the three groups made progress on the comprehensibility dimension. The means of all three groups were much higher at the end of the academic year than at the beginning of the academic year. MANOVA revealed significant group differences at T2 [$F(2, 99) = 6.909$; $p = .002$] and T3 [$F(2, 99) = 4.814$; $p = .010$]. A post-hoc Tukey multiple-comparison revealed no significant difference between Group 3 (100% EMI) and Group 1 (50% EMI) at T1 (mean difference = 0.26, $p = .563$), T2 (mean difference = 0.37, $p = .187$) or T3 (mean difference = -0.04, $p = .969$). However, the tests revealed that Group 3 (100% EMI) was significantly better than Group 2 (75% EMI) at T1 (mean difference = 0.63, $p = .049$), T2 (mean difference = 0.82, $p = .001$) and T3 (mean difference = 0.45, $p = .038$). These results reveal that the dimension of comprehensibility was affected by intensity of instruction (i.e., 100% EMI) to some extent.

Table 4 includes the descriptive statistics of average score on the content dimension. Figure 4 further illustrates the descriptive findings from Table 4.

Table 4. Descriptive statistics of content performance

	T1			T2		T3	
	N	M	SD	M	SD	M	SD
Group 1 (50%)	39	1.62	0.90	2.26	0.85	2.77	0.87
Group 2 (75%)	32	1.44	0.80	2.09	1.11	2.94	1.13
Group 3 (100%)	31	2.10	1.10	2.68	0.97	2.84	0.93

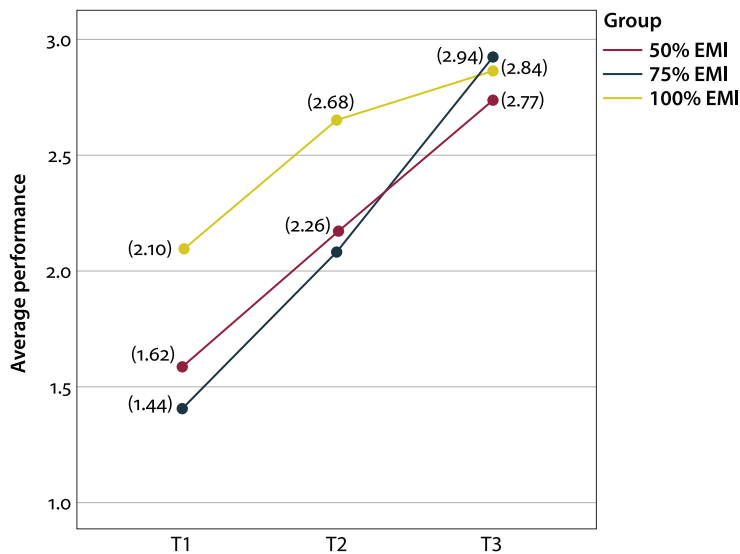


Figure 4. Content performance across groups

From Table 4 and Figure 4, we can see that the three groups improved over the academic year. The means at T3 are higher than those obtained at T1 and T2. MANOVA revealed significant group differences at T1 [$F(2,99)=4.149$; $p=.019$], but not at T2 [$F(2,99)=2.992$; $p=.055$] or T3 [$F(2,99)=0.260$; $p=.772$]. A post-hoc Tukey multiple-comparison revealed no significant difference between Group 3 (100% EMI) and Group 1 (50% EMI) at T1 (mean difference=0.48, $p=.090$). However, the tests revealed that Group 3 (100% EMI) was significantly better than Group 2 (75% EMI) only at T1 (mean difference=0.66, $p=.018$); no significant difference was found at T2 and T3. These findings suggest that the performance of content was not generally affected by intensity of instruction over the academic year.

As for the last dimension of FA, Table 5 shows the descriptive statistics of average score on the task requirements dimension experienced by the three groups. Figure 5 further illustrates the descriptive findings from Table 5.

Table 5. Descriptive statistics of task requirements performance

	T1			T2		T3	
	N	M	SD	M	SD	M	SD
Group 1 (50%)	39	1.97	1.40	3.82	1.23	4.13	1.03
Group 2 (75%)	32	2.19	1.63	4.00	1.01	4.31	1.14
Group 3 (100%)	31	3.16	1.44	4.29	1.10	4.65	0.95

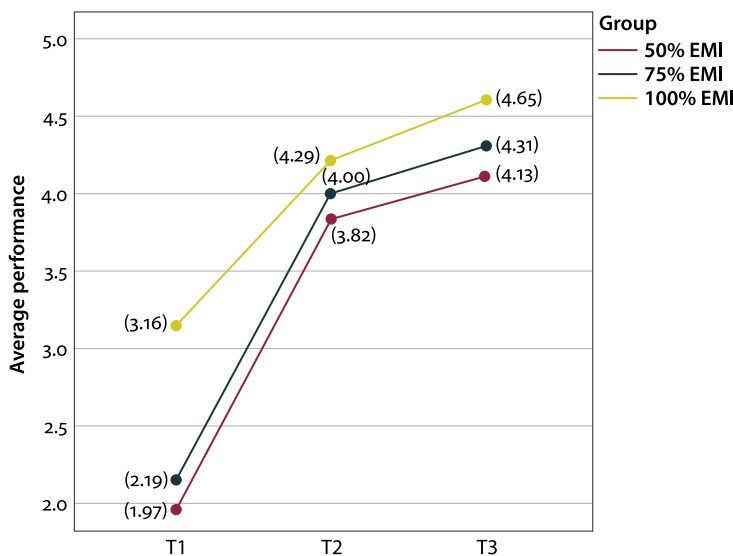


Figure 5. Task requirements performance across groups

Table 5 and Figure 5 show that the three groups improved their ability to meet task requirements over the academic year. MANOVA revealed significant group differences at T1 [$F(2, 99) = 5.956$; $p = .004$], but not at T2 [$F(2, 99) = 1.503$; $p = .227$] or T3 [$F(2, 99) = 2.129$; $p = .124$]. A post-hoc Tukey multiple-comparison revealed that Group 3 (100% EMI) was significantly better than Group 1 (50% EMI) (mean difference = 1.19 $p = .004$) and Group 2 (75% EMI) (mean difference = 0.97, $p = .029$) at T1. No significant group difference was found at T2 and T3. These findings indicate the benefit of intensity of instruction (i.e., 100% EMI) was found only at the beginning of the academic year.

5. Discussion

The present study intended to contribute to the current literature on pragmatic development in an EMI context by assessing L2 English learners' functional adequacy (FA) in written production over one academic year. Results revealed effects of intensity of exposure on gains in FA. More specifically, students with greater exposure to English (i.e., EMI 100% group) performed better in the coherence and comprehensibility dimensions of their writing at T1, T2 and T3 than those with less exposure to English (i.e., EMI 75%). These students (i.e., EMI 100%) were also significantly better than the group with less exposure (i.e., EMI 75%) in cohesion (T1 and T2), content (T1), and task requirements (T1). Finally, students in Group 3

(i.e., 100% EMI) outperformed the group with the lowest exposure to English (i.e., EMI 50%) in coherence at T2 and task requirements at T1.

Overall superiority of Group 3 (100% EMI)'s performance on the dimensions of coherence and comprehensibility can be explained by their academic experiences. Throughout the academic year, Group 3 students (100% EMI) were exposed to oral and written English on a daily basis via course materials and communication on virtual platforms (e.g., online forums and email communication). Also, students in Group 3 (100% EMI) were asked to complete a wide range of assignment types during their academic year at university. In this regard, they were required to practice writing for different genres (i.e., laboratory reports, analyses, self-learning diaries, research proposals and literature reviews). These varied opportunities for English use probably helped them improve their ability to write in a comprehensible, coherent, and cohesive manner. These findings indicate that intensity of EMI instruction could contribute to learners' gains in functional adequacy to some extent.

Despite showing the greatest performance in the content dimension at T1, students in Group 3 (100% EMI) did not perform better in this dimension at T2 and T3. We can interpret the findings based on the definition of EMI. Typically EMI refers to teaching university-level academic courses through the medium of English. Although EMI may facilitate incidental learning of dimensions specific to academic writing such as coherence, cohesion and comprehensibility through exposure to academic courses in general, content performance may involve domain-specific literacy that can be improved through courses specific to academic disciplines (Herraz-Martínez & Alcón-Soler, 2018). These results are in line with Salaberri and Sánchez-Pérez (2015), who highlight the need to incorporate new methodologies that help EMI students integrate their writing abilities with the content of the courses.

When it comes to group differences between the groups with the highest and lowest exposure to English (i.e., 100% EMI and 50% EMI), Group 3 (100% EMI) only outperformed Group 1 (50% EMI) in coherence at T2 and task requirements at T1. Taking into account exposure to English, we would have expected significant group differences between these two groups in all dimensions over the academic year. Nonetheless, our findings did not confirm the advantage of full immersion programs (i.e., 100% EMI). On the contrary, we found that Group 1 (50% EMI)'s average scores were very similar to Group 3 (100% EMI)'s on the dimensions of coherence (T1, T2 and T3), cohesion (T1 and T3), comprehensibility (T1, T2 and even higher at T3) and content (T3). Put together, our findings seem to suggest that other variables than exposure to English may affect pragmatic development (functional adequacy) in the EMI context. A possible explanation might be related to the effect of teaching strategies on learners' FA improvement during

the academic year. This makes sense if we consider the importance of feedback in improving L2 pragmatic knowledge. Accordingly, students in Group 1 (50% EMI) might have been exposed to some teaching strategies that facilitated their learning process. The lack of exposure to English (only 50% EMI) may have been compensated by the positive effects of teaching strategies. In this regard future studies may explore the effect of teaching approaches on FA development in EMI contexts.

In terms of the long-term effect of EMI, results revealed that only the coherence and comprehensibility dimensions were affected by intensity of instruction throughout the academic year (T₁, T₂ and T₃). In other words, other EMI groups (50% and 75% EMI) made similar improvement in the dimensions of cohesion, content and task requirements. These findings suggest that students may have undergone a process of academic adaptation. In addition, students may have also experienced the need to reconstruct and renegotiate their identities to fit into an EMI environment, which implies extra academic stress (Huizhu, 2012). In other words, students were so concerned about meeting the EMI academic goals and requirements during the first semester (T₁–T₂) that they may have put a lot of energy and efforts to achieve good academic performance. Accordingly, once they had made the necessary adjustments to adapt to the new setting (Gu, 2009), they did not feel academic pressure and stress and they were familiarized with the requirements to achieve good academic performance (i.e., linguistic and content-related performance) throughout the academic year. In this regard, future research needs to consider the effect of academic adaptation on FA development in EMI settings.

6. Conclusion, limitations and pedagogical implications

This study went beyond previous research conducted in EMI settings by tracing pragmatic development (functional adequacy) in three different learner groups over time. It was the first longitudinal study analyzing the effect of intensity of exposure to EMI on pragmatic development. Results showed that learners' development of functional adequacy in writing was affected by intensity of exposure to English. More specifically, the performance of five dimensions (coherence, cohesion, comprehensibility, content and task requirements) was affected positively by full immersion programs (i.e., 100% EMI). However, the effect of intensity of EMI was not all-encompassing because it was not maintained long term in the dimensions of cohesion, content and task requirements.

This study has several limitations that can be addressed in future research. First, the present study collected three writing samples from the same students over time using the identical task. Following the TBLT framework proposed by

Ellis (2003), the data collection process simulated a real-world task in which learners requested an opportunity to conduct an internship in a company related to their field of expertise. To this end, the structure of the task, topic, and type of text were kept constant over the academic year. However, we acknowledge that the effect of task repetition may have influenced the results. In this regard, future research needs to consider the effect of task familiarity on pragmatic development.

The second limitation is related to the raters who evaluated the writing samples. Although three raters (the first author and two external raters) completed norming sessions and interrater reliability was acceptable, there is always a risk of subjectivity and rater variation affecting the findings. Future research should consider this aspect by providing more rigorous training.

The third limitation is related to lack of data describing the nature of target language input students received in the EMI programs. Target language input may differ greatly corresponding to teachers' proficiency levels and instruction styles. Future research could explore input-related variables further to see whether they have any effect on learners' development of functional adequacy.

Finally, our study offers several pedagogical implications. First, the study revealed that full immersion programs (i.e., 100% EMI) promoted students' development of functional adequacy after a year. In this regard, the EMI setting seems to offer an optimal environment for the development of functional adequacy in L2 English. On the one hand, EMI programs promote effective learning of English as there are opportunities to use language both in social and academic settings. On the other hand, English is used to perform real-life tasks on an everyday basis in different communicative situations – lectures, discussion, presentations – that involve learners' ability to convey meaning effectively, i.e., functional adequacy. This study also revealed the importance of exposure to different genres and academic tasks for promoting pragmatic development in EMI programs. Given that academic writing takes many forms and there are a number of writing conventions to learn, it seems advisable to make teachers aware of the importance of practicing writing for different genres and task types (i.e., essays, laboratory reports, dissertations). Hence, teacher training in EMI settings is an important consideration in order to ensure that students receive adequate exposure to academic language use.

References

- Ament, J. R., & Pérez-Vidal, C. (2015). Linguistic outcomes of English medium instruction programmes in higher education: A study on economics undergraduates at a Catalan university. *Higher Learning Research Communications*, 5(1), 47–68.
<https://doi.org/10.18870/hlrc.v5i1.239>
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, 63(1), 68–86.
<https://doi.org/10.1111/j.1467-9922.2012.00738.x>
- Burmeister, P., & Daniel, A. (2002). How effective is late partial immersion? Some findings from a secondary school program in Germany. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 499–516). Trier, Germany: Wissenschaftlicher Verlag Trier.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
- De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012a). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121–142). Amsterdam, The Netherlands: John Benjamins.
<https://doi.org/10.1075/llt.32.06jon>
- De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012b). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34(1), 5–34.
<https://doi.org/10.1017/S0272263111000489>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.
- Fragai, E. (2001). La programmazione didattica: Il Glotto-Kit come strumento per valutare i liv-elli in entrata. In M. Barni & A. Villarini (Eds.), *La questione della lingua per gli immigrati stranieri. Insegnare, valutare e certificare l'Italiano L2* (pp. 191–208). Milan, IT: Franco Angeli Editore.
- Fragai, E. (2003). Valutare la competenza linguistico-comunicativa in italiano L2: Il Glotto-Kit per bambini e adolescenti stranieri. *Didattica & Classe Plurilingue*, 7, 1–5.
- Fulcher, G. (1987). Tests of oral performance: The need for data-based criteria. *ELT Journal*, 41(4), 287–291. <https://doi.org/10.1093/elt/41.4.287>
- Gomez-Laich, M. P., & Taguchi, N. (2018). Task complexity effects on interaction during a collaborative persuasive writing task: A conversation analytic perspective. In N. Taguchi & Y. Kim (Eds.), *Task-Based Approaches to Teaching and Assessing Pragmatics* (pp. 84–109). Amsterdam, The Netherlands: John Benjamins Publishing.
<https://doi.org/10.1075/tblt.10.04gom>
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantic*, Vol. 3 (pp. 41–58). New York, NY: Academic Press.
- Gu, Q. (2009). Maturity and interculturality: Chinese students' experience in UK higher education. *European Journal of Education*, 44(1), 37–51.
<https://doi.org/10.1111/j.1465-3435.2008.01369.x>

- Herraiz-Martinez, A., & Alcón-Soler, E. (2018). English-medium instruction and functional adequacy in L2 writing. In A. Sánchez-Hernández & A. Herraiz-Martinez (Eds.), *Learning second language pragmatics beyond traditional contexts* (pp. 145–170). New York, NY: Peter Lang.
- Housen, A. (2012). Time and amount of L2 contact inside and outside the school: Insights from European schools. In C. Muñoz (Ed.), *Intensive exposure experiences in second language learning* (pp. 111–140). Bristol, UK: Multilingual Matters.
<https://doi.org/10.21832/9781847698063-009>
- Huizhu, J. H. (2012). *Mutual influences between learners' identity construction and English language learning in the first year of university study in China*. Unpublished Ph.D. thesis, The University of Hong Kong, China.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275. <https://doi.org/10.1177/0265532208101008>
- Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16(2), 81–96.
<https://doi.org/10.1016/j.asw.2011.02.003>
- Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing*, 34(3), 321–336. <https://doi.org/10.1177/0265532216663991>
- Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 81–100). Eurosla Monographs Series 1.
- Lightbown, P. M., & Spada, N. (1991). Étude des effets à long terme de l'apprentissage intensif de l'anglais, langue seconde, au primaire. *The Canadian Modern Language Review*, 48(1), 90–117. <https://doi.org/10.3138/cmlr.48.1.90>
- Loranc-Paszylk, B. (2007). Evaluation of foreign language achievement based upon a CLIL programme in tertiary education: A Polish perspective. In R. Wilkinson & V. Zegers (Eds.), *Integrating content and language in higher education*. Maastrich, The Netherlands: Maastricht University Language Centre.
- Martin-Laguna, S. (2018). Learning pragmatics in the multilingual classroom: Exploring multicompetence across types of discourse-pragmatic markers. In A. Sánchez-Hernández & A. Herraiz-Martínez (Eds.) *Learning second language pragmatics beyond traditional contexts* (pp. 205–223). New York, NY: Peter Lang.
- McNamara, T., & Roever, C. (2007). *Language testing: The social dimension*. Oxford, UK: Blackwell. <https://doi.org/10.2307/j.ctt1ckpcf.13>
- Ohta, A. S. (2001). A longitudinal study of the development of expression of alignment in Japanese as a foreign language. In G. Kasper & K. Rose (Eds.), *Pragmatics and language teaching* (pp. 103–120). Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9781139524797.010>
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590–601. <https://doi.org/10.1093/applin/amp045>
- Revesz, A., Ekiert, M., & Torgersen, E. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37(6), 828–848.
- Salaberri, M. S., & Sánchez-Pérez, M. M. (2015). Analyzing writing in English-medium instruction at university. *Linguarum Arena*, 6, 45–58.
- Sánchez-Hernández, A. (2017). *Acculturation and acquisition of pragmatic routines in the study abroad context* (Unpublished Ph.D. thesis). Universitat Jaume I, Castelló, Spain.

- Sato, T. (2012). The contribution of test-takers' speech content to scores on an English oral proficiency test. *Language Testing*, 29(2), 223–241. <https://doi.org/10.1177/0265532211421162>
- Schmidt, R. (1993). Consciousness, learning, and interlanguage pragmatics. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 21–42). New York, NY: Oxford University Press.
- Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System*, 35(3), 305–321. <https://doi.org/10.1016/j.system.2007.02.001>
- Sheela, S. K., & Ravikumar, K. (2016). The importance of exposure in learning English as a second language: Strategies to be employed to improve the students' language exposure in the context of rapid changes in the field of technology. *Research Journal of English Language and Literature (RJELAL)*, 4(2), 770–774.
- Taguchi, N. (2012). *Context, individual differences, and pragmatic competence*. New York/Bristol: Multilingual Matters. <https://doi.org/10.21832/9781847696106>
- Taguchi, N. (2015). 'Contextually' speaking: A survey of pragmatic learning abroad, in class, and online. *System*, 48, 3–20. <https://doi.org/10.1016/j.system.2014.09.001>
- Taguchi, N., Naganuma, N., & Budding, C. (2015). Does instruction alter the naturalistic pattern of pragmatic development? A case of request speech act. *TESL-EJ*, 19(3), 1–25.
- UCLES (University of Cambridge Local Examinations Syndicate). (2001). *Quick placement test*. Oxford, UK: Oxford University Press.
- Upshur, J. A., & Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1), 3–12. <https://doi.org/10.1093/elt/49.1.3>
- Usó-Juan, E., & Martínez-Flor, A. (2006). Approaches to language learning and teaching: Towards acquiring communicative competence through the four skills. In E. Usó-Juan & A. Martínez-Flor (Eds.), *Current trends in the development and teaching of the four language skills* (pp. 3–26). Berlin, Germany: Mouton de Gruyter. <https://doi.org/10.1515/9783110197778.1.3>
- Wong, R. (2010). The effectiveness of using English as the sole medium of instruction in English classes: Student responses and improved English proficiency. *Porta Linguarum*, 13, 119–130.
- Zhang, L. J., & Yan, R. (2012). Impact of immersion teaching on English sociopragmatic awareness of Chinese kindergarten children: A polite study. *International Education*, 41(2), 33–45.

Appendix. Rating scale employed in the present study (adapted from Kuiken and Vedder, 2017)

Coherence: Is the text coherent (e.g. organization, structure, topic, structure)?

0	1	2	3	4	5
The text is not at all coherent. Unrelated progressions and coherence break are very common. The writer does not use any anaphoric device. It does not have global sense.	The text is scarcely coherent. The writer often uses unrelated progressions; when coherence is achieved it is often done through repetitions. Only a few anaphoric references are used. There are some coherence breaks. It is really difficult to reconstruct its global sense.	The text is somewhat coherent. Unrelated progressions and/or repetitions are frequent. More than two sentences in a row can have the same subject (even if the subject is understood). Some anaphoric devices are used. There can be a few coherence breaks. It is possible to reconstruct its global sense with some difficulty.	The text is coherent. Unrelated progressions are somewhat rare, but the writer sometimes replies on repetitions to achieve coherence. A sufficient number of anaphoric devices is used. There may be some coherence breaks. The text has global sense.	The text is very coherent: when the writer introduces a new topic, it is usually done by using connectives phrases. Repetitions are very infrequent. Anaphoric devices are numerous. There are no coherence breaks. Its global sense can easily be appreciated.	The writer ensures extreme coherence by integrating new ideas in the text with connectives or connective phrases. Anaphoric devices are used regularly. There are a few incidences of unrelated progressions and no coherence breaks. The text has a complete and total global sense.

Cohesion: Is the text cohesive (e.g. conjunctions, linking chunks, verbal constructions)?

0	1	2	3	4	5
The text is <i>not at all cohesive</i> . Connectives are hardly ever used and ideas are unrelated.	The text is <i>not very cohesive</i> . Ideas are not linked by connectives, which are rarely used.	The text is <i>somewhat cohesive</i> . Some connectives are used, but they are mostly conjunctions.	The text is <i>cohesive</i> . The writer makes good use of connectives, sometimes not limiting this to conjunctions.	The text is <i>very cohesive</i> and ideas are well linked by adverbial and/or verbal connectives.	The structure of the text is <i>extremely cohesive</i> thanks to a skillful use of connectives (especially linking chunks, verbal constructions and adverbials), often used to describe relationships between ideas.

Comprehensibility: How much effort is required to understand text purpose and ideas?

0	1	2	3	4	5
The text is not at <i>all</i> <i>comprehen-sible</i> . Ideas and purposes are unclearly stated and the efforts of the reader to understand the text are ineffective.	The text is <i>scarcely</i> <i>comprehen-sible</i> . Its purposes are not clearly stated and the reader struggles to understand the ideas of the writer. The reader has to guess most of the ideas and purposes.	The text is <i>somewhat</i> <i>comprehen-sible</i> . Some sentences are hard to understand at a first reading. A second reading helps to clarify the purposes of the text and the ideas conveyed, but some doubts persist.	The text is <i>comprehen-sible</i> . Only a few sentences are unclear but are understood, without too much effort, after a second reading.	The text is <i>easily</i> <i>comprehen-sible</i> and reads smoothly. Comprehensibility is not an issue.	The text is <i>very easily</i> <i>comprehen-sible</i> and highly readable. The ideas and the purpose are clearly stated.

Task requirements: Have the tasks requirements been fulfilled successfully (e.g. genre, speech acts, register)?

0	1	2	3	4	5
<i>None of the questions</i> and the requirements of the task have been answered	<i>Some (less than half)</i> of the questions and the requirements have been answered.	<i>Approximately half</i> of the questions and requirements of the task have been answered.	<i>Most (more than half)</i> of the questions and the requirements of the task have been answered.	<i>Almost all</i> the questions and the requirements of the task have been answered.	<i>All</i> the questions and the requirements of the task have been answered.

Content: Is the number of information units provided in the text adequate and relevant?

0	1	2	3	4	5
The number of ideas is not at <i>all</i> <i>adequate</i> and insufficient and the ideas are unrelated to each other.	The number of ideas is <i>scarcely</i> <i>adequate</i> and the ideas lack consistency.	The number of ideas is <i>somewhat</i> <i>adequate</i> , even though they are not very consistent.	The number of ideas is <i>adequate</i> and they are sufficiently consistent.	The number of ideas is <i>very adequate</i> and they are very consistent to each other.	The number of ideas is <i>extremely</i> <i>adequate</i> and they are very consistent to each other.

Address for correspondence

Ana Herraiz-Martinez
Universitat Jaume I
Department of English Studies
Vicent Sos Baynat
Castellón 12071
Spain
aherraiz@uji.es

Co-author information

Eva Alcón-Soler
Department of English Studies
Universitat Jaume I
alcon@uji.es