Issue 45, Enero 2026 27-42

Development and validation of a questionnaire to assess students' perceptions of EMI degree programmes

Anna szczesniak (corresponding author)
Elvira barrios
University of Málaga

Received: 2024-12-17 / Accepted: 2025-10-17
https://doi.org/10.30827/portalin.vi45.32331
Porta Linguarum ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: Given the growing popularity of English-taught degree programmes in higher education and the need to evaluate their effectiveness, this study aimed to develop and validate a questionnaire to assess students' perceptions of such programmes. Specifically, four scales were constructed to measure perceptions of students' language development, academic achievement, use of English in class, and teaching quality. The questionnaire was administered to a sample of 256 students enrolled in different English-taught degree programmes at Spanish universities. Confirmatory factor analysis (CFA) was employed to assess the psychometric properties of the scales. After refinement, the four scales demonstrated strong fit indices and solid discriminant validity. Unlike previous studies that often lack explicit reporting of instrument validity, our research provides a reliable and valid tool for researchers to explore students' perceptions of EMI, for lecturers to improve their teaching practices, and for policymakers to inform and evaluate educational reforms.
Keywords: EMI, English Medium Instruction, questionnaire, validation, student perceptions

Desarrollo y validación de un cuestionario para evaluar las percepciones de los estudiantes sobre los programas de grado impartidos en inglés

RESUMEN: Dada la creciente popularidad de los programas de grado impartidos en inglés en la educación superior y la necesidad de evaluar su eficacia, este estudio tuvo como objetivo desarrollar y validar un cuestionario para evaluar las percepciones de los estudiantes sobre dichos programas. Específicamente, se construyeron cuatro escalas para medir las percepciones de los estudiantes sobre el desarrollo del idioma, el rendimiento académico, el uso del inglés en clase y la calidad de la enseñanza. El cuestionario se aplicó a una muestra de 256 estudiantes matriculados en diferentes programas de grado impartidos en inglés en universidades españolas. Se empleó un análisis factorial confirmatorio para evaluar las propiedades psicométricas de las escalas. Tras un proceso de refinamiento, las cuatro escalas demostraron índices de ajuste sólidos y una válida discriminación entre constructos. A diferencia de estudios previos que a menudo carecen de un reporte explícito sobre la validez de los instrumentos, nuestra investigación proporciona una herramienta confiable y válida para que los investigadores exploren las percepciones de los estudiantes sobre estos programas, los docentes mejoren sus prácticas de enseñanza y los responsables de políticas informen y evalúen reformas educativas.
Palabras clave: EMI, instrucción en inglés, cuestionario, validación, percepciones estudiantiles

1. Introduction

The expanding use of EMI has led to a substantial growth in academic research (Macaro, 2019) with contributions from various global regions, including Europe, the Middle East, and Asia (Curle & Derakhshan, 2021) As highlighted by Macaro et al. (2018) in their analysis of 83 publications, "EMI in HE research is dominated by research questions relating to teacher and/or student beliefs, perceptions and attitudes towards its introduction and practice" (p.64). This dominance can be attributed to the importance of understanding the experiences of the main stakeholders directly involved in the teaching and learning process. This understanding is crucial for assessing the effectiveness and challenges of EMI programmes in higher education and suggests areas of improvement. Among the most commonly used tools to collect data on learners' or teachers' perceptions and beliefs in EMI settings is the questionnaire (Curle & Derakhshan, 2021), which, as Dörnyei (2003) notes, is the second most widely used instrument in applied linguistics, following language proficiency tests. Mackey and Gass (2005) indicated that questionnaires are widely favoured by researchers due to their numerous benefits, including practicality, affordability, efficiency in terms of time, flexibility, and ease of design. In the same line, Curle and Derakhshan (2021) remark that "questionnaires are very efficient in terms of the amount of effort, time, energy, and financial resources demanded of researchers in deploying and processing them" (pp. 4-5). Additionally, they highlight that this tool's non-intrusive nature encourages respondents to openly share their feelings, thoughts, and opinions, often leading to honest and direct answers without hesitation.

Nevertheless, despite these advantages, Curle and Derakhshan (2021) raised an important concern: many EMI researchers overlook key principles of questionnaire design, which can affect the accuracy of the data collected. Additionally, they pointed out that researchers in the field seem to have a limited understanding and awareness of the principles and theory involved in questionnaire design. This concern is also echoed by Macaro et al. (2018), who noted that only a few researchers conduct pilot studies, provide Cronbach's alpha values, or consult experts for feedback on questionnaire items. In this context, Dörnyei (2007) points to the importance of creating well-designed questionnaires which require researchers to follow a step-by-step process including: generating an item pool, conducting initial and final piloting, performing item analysis, and carrying out post hoc analysis. These steps are crucial as "the problem of producing invalid and unreliable data from a questionnaire that is ill-constructed is inevitable" (Curle & Derakhshan, 2021, p. 6). The importance of using valid and reliable tools is particularly crucial in a field like EMI, which has become a prominent area of study in the 21st century (Macaro, 2019), mainly due to the recent push for internationalisation in higher education (Curle & Derakhshan, 2021). The lack of well-designed research instruments might pose a challenge to obtaining accurate data, thereby affecting the quality and credibility of EMI research findings.

These concerns are significant because research on students' perceptions of English-taught programmes often depends on these tools, yet they often lack the necessary rigour to guarantee valid outcomes. For instance, Kırkgöz's (2009) study aimed to investigate first-year undergraduate students' views at a Turkish university on the academic tasks they were required to complete in English-medium classes, as well as the challenges they faced in

completing these tasks. Though the questionnaire employed was based on previous needs analyses, it lacked details on its design and validation process and did not report information on its validity and reliability. Similarly, Khan (2013), who investigated the views of postgraduate students and university professors regarding EMI policies and practices at two universities in Pakistan, developed questionnaires but failed to report reliability measures, despite explaining the instrument's development process. Meanwhile, Kym and Kym (2014), whose study aimed to examine students' perceptions regarding EMI in a university context in Korea, developed a questionnaire based on previous research, focusing on satisfaction levels, comprehension ability, and general perceptions of EMI classes, and students' needs. Although they reported high Cronbach alpha values for their scales, there is no information provided on the validation process. Belhiah and Elhami (2015), whose study explored students' and teachers' perceptions of the effectiveness of using English to teach subject matter in EMI programmes, also developed a questionnaire with feedback from colleagues and experts and then piloted it. Yet, similarly to others, the process lacked clear reporting on the psychometric properties of the instruments, such as its reliability. In a similar vein, Mira et al. (2021) sought to explore the motivations and attitudes of teachers and students towards EMI programmes at the Technical University of Madrid, focusing on the challenges of using English as the medium of instruction in engineering education. While the study detailed the development of the questionnaire, including the types of questions and the approach to collecting data, there is no direct reference to how the instrument was tested or validated for reliability or accuracy.

In Sugimoto's (2021) study, which aimed to gain an understanding of students' views on the challenges they encountered in EMI courses and the support they needed, the author recognised the sources of some of the questions in the questionnaire used to collect data. However, there is insufficient information regarding the stages of development and the validation process of the instrument. However, as Curle and Derakhshan, (2021) suggest, when researchers use and modify an existing scale, "they should also go through a pilot study and re-examine the validity of the scale to ensure that the changes they have made to the scale did not negatively affect the scale validity" (p. 16). Moreover, the authors argue that researchers should provide a detailed account of the specific changes they have implemented. If the scale has been translated into another language, they recommend presenting the newly adapted version. A comparable scenario is observed in the study by Iwaniec and Wang (2023). While the authors primarily investigated students' motivations for enrolling in EMI courses rather than their perceptions, they acknowledged incorporating and adapting items from existing questionnaires, including elements from Curle's (2018) validated Japanese English Medium of Instruction Attitude Scale (JEMIAS). However, the authors did not explicitly report conducting an independent validation process for the adapted JEMIAS elements within their research.

In contrast to the above-mentioned examples, Curle's (2018) original study on the JEMIAS scale addressed the internal consistency and construct validity of the questionnaire. The purpose of the questionnaire was to evaluate the attitudes of university students and lecturers towards EMI within the Japanese context. Both the professor and student versions of the questionnaire underwent validation using both qualitative and quantitative methods. For the student version, internal consistency was verified through the Cronbach alpha co-

efficient, and construct validity was assessed using Principal Component Analysis (PCA), Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA). These validation procedures demonstrated that the final version of the questionnaire exhibited strong psychometric properties.

The present study emphasises the importance of using reliable tools to explore students' perceptions in such a growing area or research. The views of those directly involved in the process of learning are crucial to evaluate an EMI programme and identify its strengths and weaknesses, with the ultimate goal of improving students' overall learning experience. Therefore, ensuring the reliability of the instruments used to collect data is essential to strengthen the credibility of the findings. To address these needs, the present study aims to develop and validate a four-scale questionnaire with robust psychometric properties to assess students' perceptions of EMI programmes at Spanish universities.

2. Methodology

2.1 Sample

A total of 256 participants took part in this study. Participants' ages ranged from 18 to 46 years, with the largest age groups being 21 (21.7%), 20 (15.7%), and 22 (15%). A total of 32.4% were men and 66% were women. The majority of them were Spanish (91.4%), while the remaining participants came from various countries across different continents, including Europe (e.g., Austria, Belgium, Czech Republic, Greece, Hungary, Slovakia), Latin America (e.g., Argentina, Brazil, Chile, Cuba, Mexico, Venezuela), North America (e.g., Canada, United States), and Asia (e.g., China, Kazakhstan, Pakistan). The sample reflected a wide range of academic backgrounds, with participants enrolled in various degree programmes, such as Administration and Business, Education, Physical Activity and Sports, Biomedicine, Engineering, Communication and Media Studies, Economics, and other specialised programmes, including Master's degree programmes. Participants came from a variety of universities, with the largest samples coming from institutions such as the University of Málaga, University Carlos III (Madrid), and the University of Granada. The largest group of participants were in their second year of study (47.3%), followed by those in their fourth year (27.7%). The sample also included students with a wide range of self-reported language proficiency levels. The highest percentage (48%) reported a B2 upper-intermediate level, followed by 23.0% at the C1 advanced level and 15.6% at the B1 intermediate level. Smaller proportions included 8.2% of students without an accredited language level and 4.3% with self-reported C2 proficiency.

2.2. Instrument

An online questionnaire was designed to assess students' perceptions of EMI programmes. The first section of the instrument aimed to collect background information about the students, including: age, gender, degree programme, university, year of study, accredited level of English or perceived level of English. The second part consisted of four scales measuring perceptions of: (1) academic achievement (6 items), (2) English language development (9

items), (3) English language use in class (6 items), and (4) teaching quality (8 items). Each statement in the questionnaire was assessed using a 10-point Likert-type scale, ranging from 0 ("Strongly disagree") to 10 ("Strongly agree").

The development of the scales followed a series of steps. First, a detailed literature review was conducted using bibliographic databases such as Web of Science and Scopus. This review aimed to identify existing research that employed questionnaires to assess students' perceptions of EMI programmes. By examining previous studies, this stage provided a foundation for understanding the types of questions used and the aspects explored in similar research contexts. It was then decided that perceptions of four dimensions would be measured: academic achievement, English language development, English language use in class, and teaching quality.

Following the literature review, a set of items was generated by two experts in English language teaching pedagogy. These items were carefully selected to ensure that they align with the aim of the study and the dimensions being measured. A significant portion of the items was either adapted or directly taken from the questionnaire developed by Barrios et al. (2022). Specifically, in Scale 2 (Language Development), Items 3, 4, 5, 6, and 7 were adapted and used in the present study. In Scale 3 (Use of English in Class), Items 2, 3, 4, 5, and 6 were taken from the same source, focusing on students' use of English in academic settings. In Scale 4 (Teaching Quality), Items 2, 3, and 4 were also drawn from Barrios et al. (2022). In Scale 1 (Academic performance), Items 2, 3, and 4 were adapted from a questionnaire on students' expectations towards EMI generated by the authors, while Items 5 and 6 were taken and adapted for the EMI context from Satayev et al. (2022). The remaining items across all scales were generated by the researchers to ensure adequate coverage of the relevant dimensions. Subsequently, a draft version of the survey in Spanish was submitted for review by a language assessment specialist. This specialist's role was to assess the clarity and appropriateness of the language used in the scales to ensure that it was easily understandable and free from ambiguity. The feedback provided by the language specialist was then incorporated into the draft. Following these revisions, the final version of the survey was completed and prepared for data collection.

Data was gathered from June to July 2024. A pilot study was performed with an in-person group of 55 students enrolled in the Bachelor's degree programme in Physical Activity and Sport at the University of Málaga. After completing the questionnaire, the students were invited to provide any questions, concerns, or suggestions. As no further modifications were proposed, this version became the final one, and the pilot study participants were included in the final sample. The authors first consulted various sources, particularly the website https://www.unportal.net/es/universitat/graus_angles/, which provides a list of English-taught programmes taught. After verifying this information through the official websites of the respective universities, emails were sent to programme coordinators requesting their collaboration. These emails included a link to the online questionnaire.

2.3. Statistical analyses

To analyse the psychometric properties of the scales, the initial step involved calculating descriptive statistics for each item, including means, standard deviations, skewness,

and kurtosis. Multivariate normality was also assessed by calculating Mardia's multivariate coefficient. Subsequently, Confirmatory Factor Analyses (CFAs) were conducted to examine the construct validity of the four scales. This study treated the variances of error terms as free parameters and fixed one structural coefficient for each latent variable (factor) to one, ensuring its scale matched that of a corresponding observable variable (item). The analysis used the maximum likelihood estimation method, enhanced with bootstrap resampling to address non-normality, particularly when skewness and kurtosis values fell within tolerable ranges (Byrne, 2013; Kline, 2011). In line with Thompson's (2004) guidance for CFA, the focus extended beyond assessing the fit of the original theoretical model to include comparisons of fit indices across multiple alternative models to determine the best fit. Based on modification indices, adjustments were made by removing items that poorly fit the model or introducing covariances between items within the same factor.

To evaluate the model's goodness of fit, several fit indices were calculated. The chi-squared ($\chi 2$), Goodness of Fit Index (GFI), and Root Mean Square Error of Approximation (RMSEA) were used as measures of absolute fit. The Adjusted Goodness of Fit Index (AGFI), Tucker-Lewis Index (TLI), and Comparative Fit Index (CFI) served as incremental fit indices. Additionally, the chi-squared divided by degrees of freedom (CMIN/DF) and the Akaike Information Criterion (AIC) were used to assess parsimonious fit (Gelabert et al., 2011). According to existing literature, the following ranges are considered indicative of good or excellent fit: $\chi 2$ p-value < 0.05; GFI/CFI/TLI 0.90-0.94/0.95-1.00; RMSEA 0.06-0.10/0.00-0.05; and CMIN/DF should fall within 0.00-3.00/3.01-5.00 (Hu & Bentler, 1999; Ruiz et al., 2010).

Convergent validity was assessed by examining the item saturations on each factor and calculating the average variance extracted (AVE). Values of ≥ 0.70 for item saturations and ≥ 0.50 for AVE were considered satisfactory (Fornell & Larcker, 1981). To evaluate the reliability (internal consistency) of each scale, Cronbach's alpha (α) (Nunnally & Bernstein, 1994) and Omega coefficients (ω) (Revelle & Zinbarg, 2009) were calculated. According to previous research, Cronbach's alpha and Omega values are interpreted as follows: acceptable = 0.70-0.79, good = 0.80-0.89, and excellent = 0.90-1.00 (Nunnally & Bernstein, 1994; Rial et al., 2006). All statistical analyses were conducted using SPSS version 25.0 and AMOS version 24.0 for Windows.

3. Results

3.1. Descriptive analyses

Table 1 summarises the results of descriptive analyses for the four scales on students' perceptions. The skewness and kurtosis values suggested that the distribution was normal for all items (except kurtosis for item 1 of scale 2), as they were both below 2.0 and 3.0, respectively (Kline, 2011). Moreover, Mardia's multivariate coefficients were above 70, so multivariate normality could be inferred (Rodríguez & Ruiz, 2008).

Table 1. *Descriptive analysis of all items of the four examined scales*

ITEM[a]	MEAN	SD	SKEWNESS	KURTOSIS
Scale 1: Perceptions of academic performance				
1. I believe that studying this degree in English has been beneficial for my academic training.	7.07	2.61	-0.96	0.35
2. I believe that studying this degree in English has significantly contributed to my understanding of the content in my field of study.	6.59	2.74	-0.69	-0.28
3. I believe that studying this degree in English has significantly contributed to the development of my practical skills in my field of study.	6.50	2.62	-0.69	-0.08
4. I believe that the practical assignments in English have provided me with skills applicable to real-life work situations.	6.43	2.71	-0.64	-0.38
5. I believe I would have learned more about the contents of the degree if the classes had been taught in Spanish.	5.56	3.40	-0.26	-1.21
6. I would have preferred to receive the classes in this degree in Spanish.	4.42	3.54	-0.21	-1.28
Mardia's multivariate coefficient				13.85
Scale 2: Perceptions of English language development				
1. I believe that my level of English has been appropriate for comfortably following the classes.	8.36	1.96	-1.71	3.47
2. I believe that my English proficiency related to my field of study has improved as a result of my enrolment in this degree.	7.14	2.62	-0.96	0.19
3. I believe that my general English proficiency has improved as a result of my enrolment in this degree.	6.71	2.59	-0.84	0.20
4. I believe that my listening comprehension (understanding spoken English) has improved as a result of my enrolment in this degree.	6.86	2.65	-0.89	0.21
5. I believe that my reading comprehension (understanding written English texts) has improved as a result of my enrolment in this degree.	6.92	2.46	-0.84	0.46
6. I believe that my writing skills (writing texts in English) have improved as a result of my enrolment in this degree.	6.16	2.69	-0.50	-0.57
7. I believe that my ability to express myself orally in English has improved as a result of my enrolment in this degree.	6.47	2.68	-0.76	-0.16
8. I believe that my technical vocabulary in English in the specific context of my field of study has improved as a result of my enrolment in this degree.	7.43	2.27	-1.09	1.01
9. I believe that studying this course of the degree in English has increased my confidence in my language skills in general.	6.67	2.61	-0.87	0.17
Mardia's multivariate coefficient				55.79

ITEM[a]	MEAN	SD	SKEWNESS	KURTOSIS
Scale 3: Perceptions of English language use in the classroom				
1. The classes in the subjects taught in English in this degree have been mainly conducted in English.	8.01	2.13	-1.26	1.23
2. I believe that in the classes taught in English, I have had enough opportunities to speak (in oral presentations and similar activities) in English.	7.31	2.24	-1.06	1.22
3. I believe that in the classes taught in English, I have had enough opportunities to converse with the professors and my classmates in English (in dialogues, debates, etc.).	6.83	2.33	-0.78	0.39
4. I believe that in the classes taught in English, I have had enough opportunities to listen to academic English (classes, videos, etc.).	7.12	2.45	-0.88	0.24
5. I believe that in the classes taught in English, I have had enough opportunities to read academic texts and documents in English.	7.61	2.26	-1.10	0.86
6. I believe that in the classes taught in English, I have had enough opportunities to write in English (presentations, essays, portfolios, etc.).	7.30	2.33	-0.92	0.50
Mardia's multivariate coefficient				27.51
Scale 4: Perceptions of teaching quality in English				
1. I believe that the quality of teaching in English has been good.	5.92	2.58	-0.56	-0.45
2. I believe that the professors have adequate linguistic competence to teach in English.	5.88	2.83	-0.35	-0.77
3. I believe that the professors have used English correctly and appropriately.	6.16	2.70	-0.48	-0.61
4. I believe that the professors have used English with good pronunciation and understandable accents.	5.80	2.84	-0.28	-0.88
5. I believe that the professors have explained the content clearly in English.	6.15	2.53	-0.46	-0.50
6. I believe that the professors provided the necessary support to facilitate my learning of the subject through English.	6.04	2.63	-0.51	-0.46
7. I believe that the professors offered appropriate and up-to-date materials in English, both in class and/or online.	6.57	2.40	-0.67	-0.15
8. I believe that the professors have given me sufficient feedback in English (comments on my tasks and assignments) that has helped me improve my learning.	5.84	2.72	-0.42	-0.67
Mardia's multivariate coefficient				51.90

Note: SD: Standard deviation; [a] The four scales were originally administered and validated in Spanish. For the purposes of this article, they have been translated into English.

3.2. Confirmatory factor analysis

The overall CFA results of the original versions of the four scales did not show an acceptable measurement model (e.g., GFI = 0.761-0.895; RMSEA = 0.142-0.238; CFI = 0.833-0.928) (Table 2). Moreover, except for scale 4, some items were saturated below 0.70 on their predicted dimension (2 and 3 items in scales 1-2 and 3, respectively) (Table 3). However, except for scale 3 (0.477), all the AVE values were above 0.50, showing adequate convergent validity.

Subsequently, in line with Thompson's (2004) recommendation, the fit indices of various alternative models were assessed to select the most suitable one. Based on the modification index values, in each scale 2-4 items were removed (Table 3) and covariances between some items were added (scale 1: items 2 and 4; scale 2: items 5 and 6; items 6 and 7; scale 3: items 3 and 5; scale 4: items 1 and 4; items 2 and 5) (Figures 1-4) This process led to the development of the new scale 1 with 4 item, scale 2 with 5 items, scale 3 with 4 items, and scale 4 with items (See the Annex).

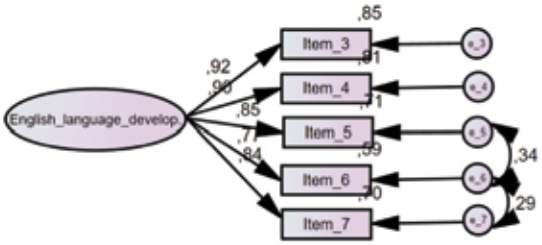

Figure 1. *Results of confirmatory factor analysis for Scale 1*

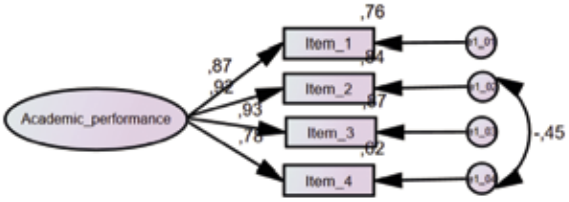

Figure 2. *Results of confirmatory factor analysis for Scale 2*

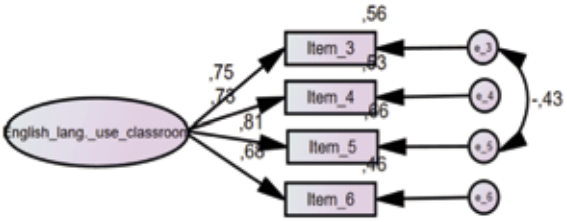

Figure 3. *Results of confirmatory factor analysis for Scale 3*

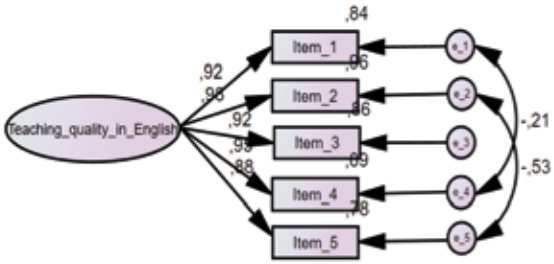

Figure 4. *Results of confirmatory factor analysis for Scale 4*

The overall CFA results of the new four scales showed excellent measurement models (e.g., GFI = 0.997-1.000; RMSEA < 0.001; CFI = 1.000) (Table 2). Moreover, except for item 6 in scale 3 (0.681), all the items were saturated above 0.70 on their predicted dimension (0.731-0.980) (Table 3). Furthermore, all the AVE values were above 0.50, showing adequate convergent validity.

Table 2. *Absolute, incremental, and parsimonious fit indices for the generated models*

	ABSOLUTE INDICES			INCREMENTAL INDICES			PARSIMONIOUS INDICES	
MODEL	χ^2	GFI	RMSEA	AGFI	TLI	CFI	CMIN/ DF	AIC
Appropriate goodness-of-fit indices	p > 0.05	≥ 0.95	< 0.05	≥ 0.95	≥ 0.95	≥ 0.95	< 3.00	-
S1-6-item	192.407*	0.837	0.238	0.620	0.722	0.833	21.379	216.407
S1-4-item	0.198	1.000	< 0.001	0.996	1.000	1.000	0.198	18.198
S2-9-item	165.839*	0.878	0.142	0.796	0.904	0.928	6.142	201.839
S2-5-item	2.141	0.997	< 0.001	0.983	1.000	1.000	0.714	26.141
S3-6-item	83.832*	0.895	0.181	0.755	0.797	0.878	9.315	107.832
S3-4-item	0.053	1.000	< 0.001	0.999	1.000	1.000	0.053	18.053
S4-8-item	260.504*	0.761	0.217	0.569	0.864	0.903	13.025	292.504
S4-5-item	0.710	0.999	< 0.001	0.994	1.000	1.000	0.237	24.710

Note: GFI: goodness of fit index; RMSEA: root mean square error of approximation; AGFI: adjusted goodness of fit index; TLI: Tucker–Lewis index; CFI: comparative fit index; CMIN/DF: chi-squared fit index over degrees of freedom; AIC: Akaike information criterion; χ^2: chi squared; S1, S2, S3, and S4: S1 (Perceptions of academic performance), S2 (Perceptions of English language development), S3 (Perceptions of English language use in the classroom), and S4 (Perceptions of teaching quality in English). * p < 0.001

Table 3. *Standardised confirmatory factor analyses solutions for the generated models*

	SCALE 1		SCALE 2		SCALE 3		SCALE 4	
	6 items	4 items	9 items	5 items	6 items	4 items	8 items	5 items
	Factor loading							
Item 1	0.880	0.874	0.266	-	0.428	-	0.921	0.916
Item 2	0.896	0.918	0.822	-	0.805	-	0.960	0.980
Item 3	0.945	0.934	0.913	0.920	0.807	0.748	0.930	0.925
Item 4	0.754	0.785	0.851	0.899	0.719	0.731	0.938	0.945
Item 5	-0.293	-	0.852	0.845	0.621	0.809	0.881	0.882
Item 6	-0.481	-	0.821	0.766	0.690	0.681	0.785	-
Item 7	x	x	0.870	0.836	x	x	0.718	-
Item 8	x	x	0.672	-	x	x	0.732	-
Item 9	x	x	0.871	-	x	x	x	x
	Average variance extracted							
	0.559	0.759	0.630	0.749	0.477	0.524	0.745	0.855

Note: Scale 1: Perceptions of academic performance; Scale 2: Perceptions of English language development; Scale 3: Perceptions of English language use in the classroom; Scale 4: Perceptions of teaching quality in English. -: item was removed from the original version; x: no item in the original version.

3.3. Reliability

Table 4 shows Cronbach's alpha and Omega coefficients for the analysed model of the four scales. The reliability (internal consistency) values of all the factors were good to excellent (α = 0.814-0.967).

Table 4. *Cronbach's alpha (α) and Omega (ω) coefficients for the generated models*

	α	ω
Scale 1-6-item	0.855	0.872
Scale 1-4-item	0.923	0.926
Scale 2-9-item	0.934	0.935
Scale 2-5-item	0.937	0.937
Scale3-6-item	0.840	0.841
Scale 3-4-item	0.814	0.815
Scale 4-8-item	0.960	0.959
Scale 4-5-item	0.967	0.967

Note: Scale 1: Perceptions of academic performance; Scale 2: Perceptions of English language development; Scale 3: Perceptions of English language use in the classroom; Scale 4: Perceptions of teaching quality in English.

4. Discussion

he aim of the current study was to develop and validate a questionnaire with robust psychometric properties to assess students' perceptions of EMI programmes at Spanish universities. CFA was used to evaluate the construct validity of the four unidimensional scales designed to assess students' perceptions of various aspects regarding EMI programmes. After removing some items, the findings of the present study showed that four examined scales had excellent model fit (See the Annex for the validated version). Moreover, the outcomes of this study showed that the four examined scales had adequate convergent validity and reliability. Although each dimension attempted to measure students' perceptions and all of them are related to students' experiences in an EMI programme, the perceptions of academic achievement, language development, language use in class, and teaching quality reflect different facets of the EMI experiences and thus were treated as separate dimensions. Yet, it is also noteworthy that the use of four separate scales allows researchers to select and focus on the individual dimensions most relevant to their research questions. While multidimensional design would require assessing the full breadth of the concept, the approach adopted in the present study offers greater flexibility.

Apart from conducting CFA and achieving excellent model fit, another strength of the present study is related to the scales' careful design process. As mentioned above, numerous studies not only lack sufficient information on the validity of their questionnaires but also fail to address other crucial stages of the design process, such as conducting pilot studies, calculating Cronbach's alpha, or seeking expert feedback (Curle & Derakhshan, 2021). In line with the study by Kahn (2013), the present study aimed to address this deficiency by providing a detailed description of how the questionnaire was designed (e.g., item development process, item pool creation, or item selection). In line with Kym and Kym (2014), our study reported Cronbach alpha values for the four scales, and like Belhiah and Elhami (2015), the questionnaire was piloted after obtaining feedback from experts.

The results of this study have important implications for research, evaluation, and policymaking. Since the topic of university internationalisation is gaining interest in academic research and practice, the need to assess its effectiveness and identify areas for improvement with valid and reliable instruments (Curle & Derakhshan, 2021) is of the utmost importance. Our scales can be used by the researchers to gain insights into students' perceptions of EMI and by lecturers to assess the effectiveness of their classes or reflect on the students' results and make the necessary changes to improve students' learning experience. Most importantly, this information is also essential for institutions and policymakers, who rely on such data to inform, evaluate, and revise educational policies. Moreover, as analysing students' perceptions has become a dominant trend in research, as highlighted by Macaro et al. (2018), the need for a valid and reliable tool is even more pressing. The available data, if collected using unvalidated instruments, could pose significant problems. Poorly constructed tools might yield inaccurate results, potentially leading to misleading conclusions about the efficacy of EMI programmes.

Researchers are increasingly discussing a "validation crisis," which arises when the validity of research instruments is neglected, leading to scepticism about entire bodies of literature (Al-Hoorie et al., 2024). As Kelley (1927) defined it, validity refers to whether

a test accurately measures what it is intended to measure. Since many tests assess abstract concepts such as attitudes or skills, verifying their accuracy is essential to ensure reliable results. Al-Hoorie et al. (2024) argue that questionable instruments undermine the credibility of findings and that this crisis is worsened by the widespread use of unreliable measurement practices. Alarmingly, many studies claiming instrument validation failed to conduct proper psychometric validation, introducing common-method bias across the empirical literature (Al-Hoorie et al., 2024). To the best of our knowledge, no existing questionnaire assessing students' experiences of EMI has undergone a detailed design and validation process. Our study successfully addressed this gap.

As for the limitations of the present study, it is important to emphasise that the questionnaire was first created, tested, and validated in Spanish. This ensured its relevance and accuracy for the intended audience, but it may limit its use for researchers aiming to replicate or expand the study in different language settings. As a result, any efforts to apply the instrument in other linguistic environments would require appropriate modifications, along with validation processes to maintain the instrument's reliability and validity.

5. Conclusions

In conclusion, this study successfully validated a questionnaire with four unidimensional scales designed to assess students' perceptions of their academic achievement, English language development, English language use in class, and teaching quality in English-taught degree programmes. All scales showed good model fit, reliability, and convergent validity, which indicates that they accurately measure different aspects of students' EMI learning experiences.

This study fills gaps in previous research by implementing a rigorous design process that included item development, pilot testing, expert feedback, and reliability analysis. By outlining these stages, the study addresses a common weakness in EMI research, where questionnaires often lack proper validation and clear methodology. Furthermore, the study demonstrates the importance of using validated instruments to avoid the "validation crisis" described in the literature, where unreliable tools can lead to misleading conclusions about the programme's effectiveness. Thus, this psychometrically reliable questionnaire improves the quality and credibility of EMI research.

Additionally, the findings have several practical implications for different stakeholders. Educators can use the scales to identify specific areas where students face challenges, such as academic performance, language development, or teaching quality, as well as recognise areas of strength. The information gathered can help them make evidence-based decisions to address the challenges and implement improvements in areas such as teaching methods, classroom practices, and curriculum design, for instance. Consequently, this can lead to better student outcomes and higher-quality EMI education. Similarly, policymakers can use this data to make decisions on EMI programme implementation to better support both teachers and students. Finally, the validated scales offer researchers a reliable tool for conducting studies on EMI programmes to assess student outcomes and evaluate the effectiveness of teaching practices. Besides, while the questionnaire was developed and validated in Spanish, it offers a strong basis for adaptation to other languages and cultural contexts. Thus, future research can refine the scales and test their applicability in different settings.

6. References

Al-Hoorie, A. H., Hiver, P., & In Nami, Y. (2024). The validation crisis in the L2 motivational self system tradition. *Studies in Second Language Acquisition*, *46*(2), 307–329. https://doi.org/10.1017/S027226312300048

Barrios, E., López-Gutiérrez, A., & Lopez-Agudo, L. A. (2022). Language-related perceptions: How do they predict student satisfaction with a partial English Medium Instruction in Higher Education? *Journal of English for Academic Purposes*, *57*(April). https://doi.org/10.1016/j.jeap.2022.101121

Belhiah, H., & Elhami, M. (2015). English as a medium of instruction in the Gulf: When students and teachers speak. *Language Policy, 14*(1), 3-23.

Byrne, B. M. (2013). *Structural equation modeling with AMOS: basic concepts, applications and programming*. Routledge.

Curle, S. (2018). Developing and validating a Japanese English medium of instruction attitude scale (JEMIAS). *The Journal of Asia TEFL, 15*(4), 1195-1221. Downloadable from: https://www.iris-database.org/iris/app/home/detail?id=york:936114&ref=search

Curle, S. M., & Derakhshan, A. (2021). Trends in using questionnaires for EMI research: Suggestions for future improvements. *Research Methods in English Medium Instruction*, *July 2021*, 32–45. https://doi.org/10.4324/9781003025115-3

Dörnyei,Z. (2003). *Questionnaires in second language research; Construction, administration, and processing*. Mahwah, N.J.: Lawrence Erlbaum.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. New York, NY: Oxford University Press.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equations models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.2307/3151312

Gelabert, E., García-Esteve, L., Martín-Santos, R., Gutiérrez, F., Torres, A., & Subirà, S. (2011). Psychometric properties of the Spanish version of the Frost Multidimensional Perfectionism Scale in women. *Psicothema, 23*(1), 133–139.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Iwaniec, J., & Wang, W. (2023). Motivations to enrol in EMI programmes in China: an exploratory study. *Applied Linguistics Review*, *14*(6), 1563–1587. https://doi.org/10.1515/applirev-2021-0180

Khan, H. I. (2013). *An investigation of two universities' postgraduate students and their teachers' perceptions of policy and practice of English medium of instruction (EMI) in Pakistani universities*. PhD thesis. University of Glasgow. http://theses.gla.ac.uk/4451/

Kırkgöz, Y. (2009). Students' and lecturers' perceptions of the effectiveness of foreign language instruction in an English-medium university in Turkey. *Teaching in Higher Education, 14*(1), 81-93.

Kline, R. B. (2011). *Principles and practice of structural equation modelling*. Guilford Press

Kelley, T. L. (1927). *Interpretation of educational measurements.* World Book Company,

Kym, I., & Kym, M. H. (2014). Students' perceptions of EMI in higher education in Korea. *The Journal of Asia TEFL, 11*(2), 35-61

Macaro, E. (2019). Exploring the role of language in English medium instruction. *International Journal of Bilingual Education and Bilingualism, 23*(3), 263-267. https://doi.org/10.1080/13670050.2019.1620678

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. https://doi.org/10.1017/S0261444817000350

Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Routledge.

Mira, S., San José Martínez, F., Hontoria, C., Sanz-Cobena, A., Rodríguez Sinobas, L., & Cuadrado, G. (2021). Attitudes of academics and students towards English-medium instruction in Engineering Studies. *European Journal of Engineering Education*, *46*(6), 1043–1057. https://doi.org/10.1080/03043797.2021.1987392

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory.* McGraw-Hill.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients Alpha, Beta, Omega and the glb: Comments on Sijtsma. *Psychometrika, 74*(1), 145–154. doi:10.1007/s11336-008-9102-z

Rial, A., Varela, J., Abalo, J., & Lévy, J. (2006). El análisis factorial confirmatorio [Confirmatory factor analysis]. In En J. Lévy Mangin (Ed.), *Modelización con Estructuras de Covarianzas en Ciencias Sociales* (pp. 119–154), Netbiblo

Rodríguez M., & Ruiz M. (2008). The reduction of skewness and kurtosis of observed variables by data transformation: Effect on factor structure. *Psicológica, 29*(2), 205-227.

Ruiz, M. A., Pardo, A., & San Martín, R. (2010). Modelos de ecuaciones estructurales [Structural Equation Models]. *Papeles del Psicólogo*, *31*(1), 34-35

Satayev, M., Barrios, E., Fernandez-Costales, A., Agaidarova, S., Izbassarova, R., & Balta, N. (2022). The effect of CLIL combined with language instruction on language learning and the role of individual and institutional factors in students' perspectives: Empirical evidence from Kazakhstan. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(10). https://doi.org/10.29333/ejmste/12425

Sugimoto, S. (2021). Students' experience of English medium instruction classes in medical school in Japan. *JALT Journal*, *43*(1), 33–60.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis. Understanding concepts and applications*. American Psychological Association

Yang, M., O'Sullivan, P. S., Irby, D. M., Chen, Z., Lin, C., & Lin, C. (2019). Challenges and adaptations in implementing an English-medium medical program: a case study in China. *BMC Medical Education, 19*(15), 1-8.

7. Appendices

Validated questionnaire in the original Spanish version

A) Percepciones sobre el rendimiento académico:

1. Considero que estudiar este Grado en inglés ha sido beneficioso para mi formación académica.
2. Considero que estudiar este Grado en inglés ha contribuido significativamente a mi comprensión de los contenidos en mi campo de estudio.
3. Considero que estudiar este Grado en inglés ha contribuido significativamente al desarrollo de mis competencias prácticas en mi campo de estudio.
4. Considero que los trabajos prácticos en inglés me han proporcionado habilidades prácticas aplicables a situaciones laborales reales.

Por favor, comparte cualquier otro comentario o experiencia que desees sobre la adquisición de contenido en este Grado:

B) Percepciones sobre la competencia lingüística en inglés:

9. Creo que mi competencia en inglés para uso general ha mejorado como consecuencia de mi matriculación en este Grado.
10. Creo que mi habilidad de comprensión auditiva (entender el inglés hablado) ha mejorado como consecuencia de mi matriculación en este Grado.
11. Creo que mi habilidad de compresión lectora (entender textos escritos en inglés) ha mejorado como consecuencia de mi matriculación en este Grado.
12. Creo que mi habilidad de expresión escrita (escribir textos en inglés) ha mejorado como consecuencia de mi matriculación en este Grado.
13. Creo que mi habilidad de expresarme oralmente en inglés ha mejorado como consecuencia de mi matriculación en este Grado.

Por favor, comparte cualquier otro comentario o experiencia sobre tu trayectoria en inglés como consecuencia de tu matriculación en este Grado (opcional):

C) Percepciones sobre el uso de inglés en clases:

18. Creo que en las clases impartidas en inglés he tenido suficientes oportunidades para conversar con los profesores y mis compañeros de clase en inglés (en diálogos, debates...).
19. Creo que en las clases impartidas en inglés he tenido suficientes oportunidades para escuchar inglés académico (clases, vídeos, etc.).
20. Creo que en las clases impartidas en inglés he tenido suficientes oportunidades para leer textos y documentos académicos en inglés.
21. Creo que en las clases impartidas en inglés he tenido suficientes oportunidades para escribir en inglés (presentaciones, ensayos, portafolios...).

Por favor, comparte cualquier otro comentario o experiencia que desees sobre el uso de inglés en este Grado (opcional)

D) Percepciones sobre la docencia en inglés:

22. Considero que la calidad de la docencia en inglés ha sido buena.
23. Considero que el profesorado tiene una competencia lingüística adecuada para enseñar en inglés.
24. Considero que el profesorado ha utilizado inglés de manera correcta y apropiada.
25. Considero que el profesorado ha empleado inglés con buena pronunciación y acentos comprensibles.
26. Considero que el profesorado explicó los contenidos claramente en inglés.

Por favor, comparte cualquier otro comentario o experiencia que desees sobre calidad de la docencia en este Grado (opcional):