

Willingness to communicate in a foreign language in the four skills. Sex and age differences

Disposición a comunicarse en un idioma extranjero en las cuatro destrezas. Diferencias de edad y sexo

Eliseo Fernández-Barrionuevo¹, Javier Villoria Prieto², Francisco Tomás González-Fernández³, y Antonio Baena-Extremera²

¹Consejería de Educación de la Junta de Andalucía, Motril, España; ²Universidad de Granada, Granada, España;

³ Universidad Pontificia Comillas, Palma de Mallorca, España.

Abstract

The object of this research was to validate the Willingness to communicate in a foreign language in the four skills in the Classroom Scale to a Spanish speaking context. Taking the study of authors (in press) as a starting point, this instrument validation will allow to obtain information about the student FLL motivation in CLIL in regular schools. This would allow researchers and teachers to have a valid instrument that provides information about the four language skills in FLL or CLIL programmes. In addition, as part of the construct validity analysis, sex and age differences are explored in order to obtain a student profile regarding Willingness to communicate in a foreign language in the four skills. The WTC-FL in the four skills adaptation was administered to a total of 3355 students from Andalusia aged from 11 to 17 years. After the translation and adaptation processes, item and internal structure analyses were conducted. Other analyses have shown this instrument to be sex-invariant and to have convergent validity. A multi-level model analysis was also conducted in order to study the construct validity, concluding with similar results from similar studies.

Keywords: Motivation; FLL; WTC-FL; validation; sex; age.

Resumen

El objetivo de este estudio ha sido la validación de la escala *Willingness to communicate in a foreign language in the four skills in the Classroom Scale* (escala de disposición a comunicarse en una lengua extranjera en las cuatro destrezas) en un contexto de hispanohablantes. Partiendo del estudio de los autores (en prensa) como punto de partida, la validación de este instrumento permitirá obtener información sobre la motivación en el aprendizaje de lenguas extranjeras en centros educativos. Todo esto permitirá tanto a investigadores como a profesores tener un instrumento que proporcione información sobre cada una de las destrezas lingüísticas en programas AICLE. Además, como parte del análisis de la validez de constructo, las diferencias en sexo y edad han sido analizadas con la finalidad de obtener un perfil para la DAC en cada una de las cuatro destrezas. El instrumento fue administrado a un total de 3355 estudiantes de Andalucía con edades comprendidas entre los 11 y 17 años. Tras la traducción y procesos de adaptación, se ejecutaron análisis de ítems y estructura interna comparando los modelos de 5 y 7 factores. Esto dio lugar a análisis CFA, fiabilidad y validez que concluyeron el mejor ajuste del modelo de 5 factores. Otros análisis han mostrado la invarianza por sexo y la validez convergente. También se realizó un análisis de modelos multiniveles para estudiar la validez de constructo obteniendo resultados similares a otros estudios.

Palabras clave: Motivación, FLL; WTC-FL; validación; sexo; edad.

Fecha de recepción: 03/04/2020

Fecha de aceptación: 08/09/2020

Correspondencia: Francisco Tomás González-Fernández, U.P. Comillas, Palma de Mallorca, España
Email: francis.gonzalez.fernandez@gmail.com

Introduction

After the boost that the Common European Framework of reference for Languages (CEFRL) meant for the communicative approach in Foreign Language Learning (FLL), the analysis of variables related with the communication has become vital. Within this context, it is crucial to study of the Willingness to Communicate (WTC) as it is the last psychological step in the preparation of the effective communication in a given language, therefore being relevant in the communication and learning processes.

In order to conceptualize this construct (MacIntyre, 1998) developed a heuristic model where all the variables that influenced the WTC were stratified in seven layers. In the peak of the pyramid were the communication behaviour, followed by the behavioural intention or the WTC. Underneath, the next five layers were represented: the situated antecedents, the motivational propensities, the affective-cognitive context and the social and individual context.

The WTC in a Foreign Language (WTC-FL), has been widely studied, being linked with variables like attitudes and motivation (Baker & MacIntyre, 2000; MacIntyre et al., 2002; Yasmina et al., 2004) factors intrinsically related to the individual such as personality, age, gender or perceived competence (MacIntyre & Charos, 1996; MacIntyre & Doucette, 2010) or contextual variables such as group size (Cao & Philp, 2006; Kang, 2005) classroom environment (Peng & Woodrow, 2010) or social support (MacIntyre et al., 2001).

Regarding the CLIL (Content and Language Integrated Learning), as the communication is one of the composing factors of the 4Cs model that would define that approach (Coyle et al., 2010) the study of the WTC-FL is essential from a theoretical and a practical perspective. The scarce studies of this variable in CLIL settings tend show an increased WTC-FL compared to standard FLL (MacIntyre & Doucette, 2010; Menezes & Garau, 2015; Pihko, 2007; Wu, 2014) although some exceptions are found (Vanderveen, 2015).

There are some proposals regarding the WTC-FL measurement instruments. The Willingness to communicate Scale (WTCS) (McCroskey, 1992; McCroskey, 1985) is one of the most extended instruments which contains three sub-scales for different receptors, and three for distinct communicative environments. Based on the Spanish context (Díaz Pinto, 2009) designed a dichotomic version (yes/no) that afterwards would be adapted for a 7-point Likert scale (Santos-Menezes, 2014). Apart from this instrument there are others such as the WTC-meter (Kamprasertwong, 2010) adapted from the Peter D MacIntyre and Gardner (1991) (version, or the Willingness to Communicate in a Foreign Language Scale (WTC-FLS) (Baghaei et al., 2012). However, the *Willingness to communicate in the classroom* developed by P. D. MacIntyre et al., 2001) (Coyle et al., 2010) is of great interest due to its contextualization in the school setting and, importantly, because it provides separated information for each of the language skills: reading, listening, speaking and writing.

As the measuring instruments for the WTC-FL in Spanish are rare, and even more the ones centred in CLIL settings, the aim of this study is to validate the *Willingness to communicate in the classroom* scale in Spanish. This would allow researchers and teachers to have a valid instrument that provides information about the four language skills in FLL or CLIL programmes. In addition, as part of the construct validity analysis, sex and age differences are explored in order to obtain a student profile regarding WTC-FL.

Materials and Methods

Participants

The sample represented the universe with a statistical confidence level of 99% and a 2% margin of error. Therefore, the sample of this study was compounded by a total of 3.355 students ($n_{\text{girls}}=1.797$ (53.6%); $n_{\text{boys}}=1.558$ (46.4%)) that coursed CLIL studies in secondary schools in Andalucía. Their ages ranged from 11 to 17 (Mean (M)=13.77; Standard deviation (SD)= 1.483), being the M age of girls 13.75 (SD=1.486) and 13.79 (SD=1.479) for boys. The course distribution was: 1051(31.3%) for CSE (Compulsory Secondary Education) 1st grade, 595 (17.7%) for CSE 2nd

grade, 571 (17.0%) for CSE 3rd grade, 681(20.3%) for CSE 4th grade, and 457 (13.6%) for 1st bachillerato (Pre-University Secondary Education).

Process of Adaptation and Translation

The process of translation of the LLOS-IEA was undertaken following the international methodological standards that the International Test Commission (ITC) recommends adapting tests and scales from one culture into another (Hambleton et al. 2005; Muñiz, 2000; Muñiz & Bartram, 2007) In order to proceed precisely, processes of direct and back translation of the items were performed (Brislin, 1970; Brislin, 1986). Following the parallel back translation procedure (Brislin, 1986) two translators independently translated one version in the target language (Spanish); later these works were re-translated into English by two professionals who were not aware of original work. The quality of the work was assessed regarding the similarity with the original version (Hambleton et al, 2005) and there were hardly no modifications as both versions were almost similar.

In the WTC in the classroom scale all the references to French as FLL were replaced to English as a FLL. There was no need to adapt this scale to the CLIL context, to be used in content subjects though in English, as the wording of the original instrument did not explicitly mention a foreign language class. Subsequently, a qualitative evaluation (content validity) of the work was undertaken by five experts (Osterlind, 1987): two in scale design and three in the construct assessed. They were provided with an items' specification table (Calabuig & Crespo, 2009; Spaan, 2006), which included the semantic definition of the construct, its components and a list of the original and adapted items. These experts judged each item's weightiness in its domain by using a scale from 1 (not at all) to 4 (absolute), and they also assessed the item's suitability and univocity. They had the opportunity to write any concern, annotation or an alternative wording of any of the items.

The items that scored mean values <2.5 in suitability (Nuviala et al., 2008) were revised according to the experts' reviews, and if four out of five experts did not classify any item within its theoretical dimension, it was readapted again so it would clearly and accurately express the theoretical dimension. The overall item concordance of comprehensibility and suitability was measured through the Intraclass Correlation Coefficient (ICC) from a Two-way mixed model, assuming an absolute agreement. The values obtained were $ICC=.780$ for item suitability, $ICC=.894$ for item weightiness and $ICC=.871$ for univocity.

The new version was administered to 55 CSE and pre-university students aged between 12 and 18 using various options of density, item separation and general formatting (Dörnyei, 2003) that conducted to minor modifications. The final version of the LLOS-IEA was obtained after an analysis of the psychometric results, and one last revision carried out by the research team.

Procedure

After the permission from the school administrators, the questionnaire was administered informing the anonymous and voluntary nature of participation. Also, this research has ethical approval. The participation took part between January and March 2016, and lasted about twenty minutes, and concerns about comprehension were attended throughout that time. According to the Declaration of Helsinki (2008), all the respondents were briefly informed about the purpose of the study and their rights as participants, apart from being given the opportunity to give up the survey at any time.

Data Analysis

First, an item and homogeneity of the scale analysis was performed, which included: each dimension Cronbach's alpha (α); and each item M, ST, corrected item-total correlation coefficient (CITC-c), correlation between the item and its dimension (CC), the Cronbach's alpha if item were deleted, Kurtosis and Skewness. For this analysis, the SPSS v. 21 for Mac OS X was used.

Afterwards, as part of an exploratory factorial analysis (EFA), an extraction method of principal components (PCA) was performed extracting a fixed number of seven factors following the structure of the original instrument.

Later, in order to assess whether the data distribution was normal, an analysis based on the Relative Multivariate Kurtosis (RMK) of PRELIS through LISREL 8.80 programme was performed. In order to confirm the dimensionalization of the scale, the factor structure of the instrument was assessed with CFA using the Weighted Least Squares (WLS) estimation method for ordinal variables in the LISREL 8.80 (Jöreskog & Sörbom, 2003). In addition to the factor structure from the original instrument, two other 5 and 7-factor models were also compared. Regarding reliability and validity, in addition to the α value, the Composite Reliability and the Average Variance Extracted (AVE) for each dimension were also calculated. Last, the convergent validity, the construct validity and the sex invariance were determined. To study the construct validity a multi-level analysis was performed. The WTC in the classroom scale factors were selected as an independent variable, and the students' sex and age were the factors of this mixed model multi-level analysis.

Results

Items' Analysis and Scale Homogeneity

The items' statistical analysis held the item-factor distribution of the original instrument. The criteria to maintain items was: CITC-c \geq .30, SD > 1, and all the possible responses used at least once (Nunnally & Bernstein, 1995). The Kurtosis and Skewness should also be close to 0 and <7 (Curran et al., 1996) (Table 1).

Table 1.

Statistical analysis of each item of WTC in the Classroom.

Scale: <i>WTC in the Classroom</i>	M	SD	CITC-c	CC	α withoutSkewness	Kurtosis item	
Speaking ($\alpha= .75$)							
1. Hablar en grupo acerca de las vacaciones de verano	2.63	1.40	.51	.66	.72	.34	-1.19
2. Hablar a tu profesor sobre los deberes	2.56	1.29	.51	.64	.72	.36	-.98
3. Un extraño entra en tu aula. ¿Cuántas ganas tendrías de tener una conversación si fuera él el que hablara primero?	2.97	1.34	.45	.61	.73	-.03	-1.14
4. Estás confuso acerca de una tarea que tienes que completar. ¿Cuántas ganas tendrías de pedir instrucciones o aclaraciones?	3.45	1.27	.37	.53	.74	-.41	-.85
5. Hablando con un amigo mientras estás esperando en una fila	2.97	1.49	.42	.59	.74	.01	-1.41
6. ¿Cuántas ganas tendrías de ser actor en una obra de teatro?	2.72	1.51	.36	.54	.75	.24	-1.39
7. Describir las reglas de tu juego favorito	3.00	1.39	.52	.66	.72	-.03	-1.23
8. Jugar un juego en la lengua extranjera. por ejemplo, Monopoly	3.44	1.43	.48	.64	.72	-.44	-1.14
Reading ($\alpha= .83$)							
1. Leer una novela	2.61	1.49	.59	.73	.80	.37	-1.28
2. Leer un artículo de un periódico	2.68	1.39	.68	.79	.78	.25	-1.20
3. Leer cartas o emails de un amigo escritas en inglés	3.20	1.40	.67	.79	.78	-.22	-1.21
4. Leer mensajes o emails escritos específicamente para ti en las que el escritor ha utilizado a propósito palabras o construcciones sencillas	3.49	1.35	.57	.71	.80	-.49	-.93
5. Leer un anuncio en un periódico para encontrar una buena oferta para una bicicleta	2.80	1.42	.50	.66	.82	.12	-1.29
6. Leer críticas de películas populares	3.09	1.46	.56	.71	.81	-.14	-1.34
Writing ($\alpha= .85$)							
1. Escribir un anuncio para vender una bici vieja	2.54	1.37	.54	.67	.84	.38	-1.09

2. Apuntar las instrucciones de tu pasatiempo favorito	2.80	1.35	.65	.76	.82	.12	-1.16
3. Escribir un texto de tus animales favoritos y de sus hábitos	3.13	1.40	.63	.74	.83	-.13	-1.23
4. Escribir una historia	3.19	1.43	.62	.74	.83	-.21	-1.26
5. Escribir una carta a un amigo	3.41	1.36	.65	.76	.82	-.43	-1.01
6. Escribir a una revista para participar en un concurso	2.94	1.41	.62	.74	.83	.00	-1.26
7. Escribir una lista de cosas que tienes que hacer mañana	2.80	1.42	.56	.69	.84	.14	-1.30
Listening ($\alpha = .80$)							
1. Escuchar las instrucciones para completar la tarea	3.67	1.26	.52	.70	.78	-.65	-.62
2. Cocinar un pastel si las instrucciones no fueran en español	3.24	1.39	.57	.75	.77	-.27	-1.16
3. Rellenar un formulario de solicitud	3.25	1.30	.62	.77	.75	-.26	-.99
4. Seguir instrucciones en inglés	3.42	1.28	.70	.82	.73	-.40	-.89
5. Entender una película en inglés	3.65	1.35	.52	.71	.78	-.63	-0.84

Nota. M = Mean; SD = Standard Deviation; CITC-c = Corrected item-total correlation coefficient; CC = Correlation between the item and its dimension.

Items from factor 1 (speaking) showed mean values between 2.56 from item 2 and 3.55 from item 4. All the SD values were >1 , and this dimension's internal consistency was satisfactory ($\alpha = .75$). All the CITC-c were $\geq .36$.

With regard the second factor (reading), the mean values ranged from 2.61 (item 1) to 3.49 (item 4). All the SD values were >1 , and this dimension's internal consistency was satisfactory ($\alpha = .83$). CITC-c were $\geq .50$.

With regard the third factor (writing), all the items presented mean values from 2.54 (item 1) to 3.41 (item 5), SD values were >1 , and this dimension's internal consistency was suitable ($\alpha = .85$). All the CITC-c were $\geq .54$.

Finally, items from fourth factor (listening) all the mean values were between 3.24 from item 2 to 3.67 from item 1. SD values were >1 , and this dimension's internal consistency was suitable ($\alpha = .80$). All the CITC-c were $\geq .52$.

Some authors such as Carretero and Pérez (2005) recommend performing a correlation study in order to guarantee each dimension's homogeneity (CC). In this work, the correlation between each item's score and its overall component's score were $CC \geq .53$.

Internal Structure Analysis

Based on the validation process of the original instrument (McIntyre et al. 2001), an EFA for the four-factor model was performed. A PCA method was conducted, requiring a minimum correlation of .40 in order to consider each item important within the factor (Stevens, 1992), the Kaiser-Meyer-Olkin (KMO) index was good (.941), and the Bartlett's sphericity test was significant ($\chi^2_{(325)} = 332271.932$, $p < .000$), concluding with the suitability of the implementation of the EFA. The results confirmed a four-factor extraction accounting for 49.914% of the total variance explained (table 2). However, the items 5 in reading, and item 4 in writing were respectively placed in writing and reading which has taken into account in the further analysis.

Table 2.

Factor saturation and communalities.

Speaking	FLS	h^2
1. Hablar en grupo acerca de las vacaciones de verano	.66	.49
2. Hablar a tu profesor sobre los deberes	.58	.43
3. Un extraño entra en tu aula. ¿Cuántas ganas tendrías de tener una conversación si fuera él el que hablara primero?	.54	.40
4. Estás confuso acerca de una tarea que tienes que completar. ¿Cuántas ganas tendrías de pedir instrucciones o aclaraciones?	.43	.32

5. Hablando con un amigo mientras estás esperando en una fila	.62	.41
6. ¿Cuántas ganas tendrías se ser actor en una obra de teatro?	.45	.34
7. Describir las reglas de tu juego favorito	.59	.48
8. Jugar un juego en la lengua extranjera, por ejemplo Monopoly	.50	.39
Reading		
1. Leer una novela	.76	.64
2. Leer un artículo de un periódico	.70	.62
3. Leer cartas o emails de un amigo escritas en inglés	.71	.67
4. Leer mensajes o emails escritos específicamente para ti en las que el escritor ha utilizado a propósito palabras y construcciones sencillas	.59	.48
5. Leer un anuncio en un periódico para encontrar una buena oferta para una bicicleta	.59	.46
6. Leer críticas de películas populares	.49	.45
Writing		
1. Escribir un anuncio para vender una bici vieja	.75	.61
2. Apuntar las instrucciones de tu pasatiempo favorito	.67	.60
3. Escribir un texto de tus animales favoritos y de sus hábitos	.57	.47
4. Escribir una historia	.40	.46
5. Escribir una carta a un amigo	.44	.49
6. Escribir a una revista para participar en un concurso	.54	.49
7. Escribir una lista de cosas que tienes que hacer mañana	.57	.45
Listening		
1. Escuchar las instrucciones para completar la tarea	.67	.51
2. Cocinar un pastel si las instrucciones no fueran en español	.66	.51
3. Rellenar un formulario de solicitud	.70	.61
4. Seguir instrucciones en inglés	.77	.69
5. Entender una película en inglés	.63	.52

S=factor saturation, h²=communalities.

Confirmatory Factor Analysis

In order to study the psychometric properties of WTC in the classroom scale original dimensionalization, structural equation modelling was performed. Different absolute and relative fitness indices were calculated (Jöreskog & Sörbom, 2003; Hair et al., 2009) such as p-value associated with Chi-square test, χ^2 and degrees of freedom ratio (df; χ^2 /df), goodness of fit index (GFI), normed fit index (NFI), non-normed fit index (NNFI), and comparative fit index (CFI). The estimated parameters were considered significant when the value associated with the t-value was higher than 1.96 ($p < 0.05$).

Firstly, RMK analysis was conducted with this scale which resulted with a Mardia-Based-Kappa value of 1.254. Test results showed that multivariate normality could not be accepted (upper limit=1.006; lower limit=.994), which implied the use of a robust estimator. Therefore, a weighted least squares (WLS) estimation method for ordinal variables in the LISREL 8.80 (Jöreskog & Sörbom, 2003) program was conducted. The polychoric correlations matrix and asymptotic covariance matrix were used as input for data analysis.

Table 3.

Fit indices for the WTC in the classroom scale.

Scale	χ^2	df	χ^2 /df	p	GFI	NFI	NNFI	CFI	RMSEA
WTC in the classroom	1978,8	291	4.95	<.000	.98	.90	.95	.91	.04

As it can be appreciated in Table 3, all the indices of the model proposed (Figure 1) presented were within the acceptable parameters.

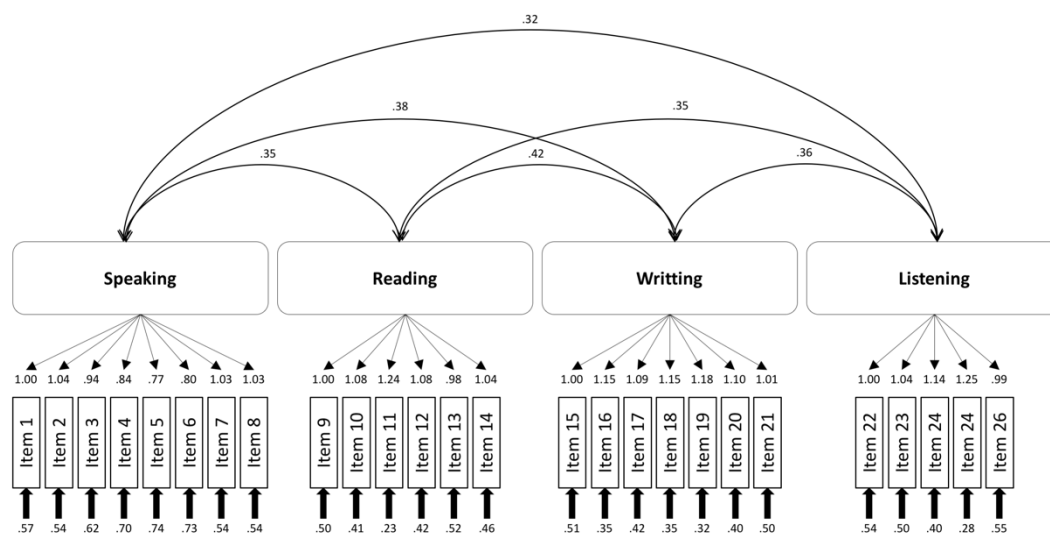


Figure 1. Path diagram of the CFA, with standardized weights and measurement errors of each one of the items in the WTC in the classroom scale.

Reliability and Validity

Table 4 shows the model 3 reliability and validity. Apart from α values, AVE (Average Variance Extracted) and composite reliability data was calculated, requiring minimum values of .70 y .50 respectively (Hair et al., 2009). As it can be appreciated in the indices of table 4, the WTC in the classroom can be considered a reliable and valid instrument.

Table 4.

Reliability and validity of the scale.

Dimension	Composite reliability	AVE	α
WTC-FL speaking	.84	.68	.75
WTC-FL reading	.89	.78	.83
WTC-FL writing	.92	.83	.85
WTC-FL listening	.85	.70	.80

Sex Invariance

In order to analyse the factorial invariance, Abalo, Lévy, Rial, and Varela (2006) recommendations were followed estimating the same model for both samples. No significant differences were found in χ^2 between models, rejecting the H0 and accepting the invariance. However, due to χ^2 sensitiveness to sample size, Cheung and Rensvold (2002) criteria regarding the ΔCFI were also implemented. According to these authors ΔCFI values $\leq .01$ indicate that the null hypothesis should not be rejected, being $\Delta CFI=.005$ in the present study. Finally, the rest of the results state that measurement properties remain sex invariant.

Construct Validity: Sex and Age Differences

A multi-level model analysis was performed in order to study the construct validity. Several models were tested considering province, school and grade, finally determining the model by school and grade as it got the best BIC (10224.732).

Table 5 displays the mixed model multi-level analysis outcome. The estimated mean values by sex and age (grouped in school cycles) adjusted to school and grade are presented. The student stipulated age for 1st cycle is 12–13 for, 14–15 for 2nd cycle and 16 for 1st bachillerato. This table also shows the standard error, the 95% confidence interval, and the statistical test corresponding to the model where the hypothesis of equal means in the dimensions between the independent variable categories is contrasted.

This table also includes the difference between answer and reference categories, and the p-value associated. In addition, this table also includes the difference between answer and reference categories, and the p-value associated to the statistical tests of margin corrected means comparison by multiple comparisons through SIDAK.

Table 5.

Mixed model multi-level analysis outcome

WTC-FL		Adjusted mean	Standard Error	95% Confidence Interval		Multilevel analysis statistical tests			SIDAK	
				Lower	Upper	Adjusted differences with the reference value	Adjusted mean	Standard Error	Lower	Upper
Speaking	Boy	2.89	.047	2.80	2.98	-.03	1.67	3351	.19	.19
	Girl	2.93	.047	2.84	3.02					
	1 st cycle	3.06	.018	2.02	30.96	.26*	37.66	3351	.00	.00
	2 nd cycle	2.79	.024	2.79	2.84	-.08				.88
	Bachillerato	2.88	.131	2.62	3.13	-.17				.44
Reading	Boy	2.82	.058	2.70	2.93	-.26*	55.94	3351	.00	.00
	Girl	3.09	.058	2.97	3.20					
	1 st cycle	2.98	.022	2.93	3.02	.035	.43	3351	.64	.73
	2 nd cycle	2.94	.030	2.88	3.00	.000				1.0
	Bachillerato	2.94	.161	2.63	3.26	-.034				.99
Writing	Boy	2.78	.057	2.67	2.89	-.22*	42.13	3351	.00	.00
	Girl	3.01	.056	2.90	3.12					
	1 st cycle	3.03	.022	2.99	3.07	.18*	13.90	3351	.00	.00
	2 nd cycle	2.84	.029	2.78	2.90	.02				.99
	Bachillerato	2.82	.156	2.51	3.12	-.21				.44
Listening	Boy	3.36	.056	3.25	3.46	-.15*	20.54	3351	.00	.00
	Girl	3.51	.055	3.40	3.62					
	1 st cycle	3.43	.021	3.39	3.47	-.013	.069	3351	.93	.97
	2 nd cycle	3.45	.029	3.39	3.50	.026				.99
	Bachillerato	3.42	.154	3.12	3.72	-.014				1.00

Concerning the dimension WTC-FL for speaking, significant differences were found in age ($p < .000$), being higher in first cycle students ($M = 3.060$; $SE = .018$; adjusted difference = $-.263$; $F = 37.66$).

Regarding the WTC-LE for listening, significant differences were found for sex ($p < .000$) as the mean values for boys were lower ($M = 2.823$; $SE = .058$, adjusted difference = $-.267$), with a very important F value ($F = 55.94$).

With respect to the dimension WTC-FL for writing, significant differences were found in both sex and age ($p < .000$). With regard sex, the value was lower in boys ($M=2.787$; $SE=.057$; adjusted difference $=-.225$; $F=42.13$), and for age, the mean values were higher in the first cycle students ($M=3.033$; $SE=.022$; adjusted difference $=.189$; $F=13.91$).

Finally, significant differences were also found for WTC-FL for listening ($p\text{-valor} < .000$), being higher the values for boys ($M=3.360$; $SE=.056$; adjusted difference $=-.154$; $F=20.54$)

Discussion

The main objective in this work has been to study and validate the Willingness to Communicate in the Classroom Scale to a Spanish speaking context. Taking the study of authors (in press) as a starting point, this instrument validation will allow us to obtain information about the student FLL motivation in CLIL and regular schools, knowing the important of motivation in physical education (Baena-Extremera, Gómez-López, Granero-Gallegos, & Martínez-Molina, 2016; Baena-Extremera, Granero-Gallegos, Sánchez-Fuentes, & Martínez-Molina, 2014; Ruiz-Juan & Baena-Extremera, 2015).

First, an item and homogeneity analysis was performed in addition to an internal structure study that did not placed two of the items (13 and 18) in the dimension proposed in the original instrument. However, after the rest of the analysis, it can be concluded, as observed in table 4 that the model gets a good fitting.

Regarding gender differences, most of the research has used the WTCS designed by McCroskey and Baer (Wu, 2014), adapted by MacIntyre y Charos (MacIntyre, 1998) which is based in the oral interaction. As the instrument performed in this study provides information for the four language skills, none of the comparisons would be totally equivalent. However, if we consider the gender differences of the WTC for the speaking skills, this study would show similar results to other investigations (Jöreskog & Sörbom, 2003; Santos Menezes, 2014), as girls performed higher mean values, though with no significant results. On the other hand, the CEFRL states that the process involving speaking interaction would integrate the speaking and listening activities in a “overlapping a cumulative” manner (p. 92). Therefore, in order to compare this investigation more precisely with other studies that analyse the gender gap in WTC for the oral interaction, it should be taken into account the listening skills besides the speaking skills. In conclusion, regarding the two oral skills (speaking and listening), girls present a higher WTC (only the listening skills shows significant values). This would be in line with other studies where girls display a higher WTC (Alemani et al., 2013; Baker & MacIntyre, 2000; Donovan & McIntire, 2004; Lin & Ranger, 2003). This gender gap might be the result of a feminized school context or a higher FLL intrinsic motivation in girls (Paper LLOS-IEA in press).

With regard to age, unlike investigations in other contexts, (Donovan & McIntire, 2004; MacIntyre et al., 1998) this study shows higher values in older students for the WTC-FL in all the skills, however, only in the case of the productive skills (writing and listening) the results were significant. In the case of the elder students, a more academic focus centred in preparing exams that hardly ever contains interaction activities would probably be one of the causes of these results. In any event, these results are relevant as it would be logical to think that the higher foreign language knowledge of senior students is the cause of increase of their WTC-FL. Therefore, it would be advisable to perform further research in order to dig into the reasons of that WTC decline through secondary education in bilingual schools and the possible effects in the FLL.

Data revealed in this study about sex and age differences would not only open new fields for future investigations, but also to provide teachers with a knowledge about their students that would allow a more individualized learning. In this manner, it would be necessary to find pedagogical strategies to increase the WTC-FL in boys and senior students who are the ones with a more reduced level of WTC-LE.

MacIntyre (MacIntyre et al., 2001) states that “perhaps the most important decision language learners face is whether or not to use their incipient skills when the opportunity arises, inside or

outside the classroom” (p. 573). However, that decision might be different depending of the language skill used, and therefore the use of one instrument that measures the WTC for each of the skills separately is highly recommended. Futures investigations that apply this instrument could reveal precise information about the communication in FLL and bilingual lessons that has been hidden to these days.

References

- Alemi, M., Tajeddin, Z., & Mesbah, Z. (2013). Willingness to communicate in L2 English: Impact of learner variables. *Research in Applied Linguistics*, 4(1), 42-61.
- Baena-Extremera, A., Gómez-López, M., Granero-Gallegos, A., & Martínez-Molina, M. (2016). Prediction model of satisfaction and enjoyment in Physical Education from the autonomy and motivational climate. *Universitas Psychologica*, 15(2), 39-49. <http://dx.doi.org/10.11144/Javeriana.upsy15-2.mpsd>
- Baena-Extremera, A., Granero-Gallegos, A., Sánchez-Fuentes, J. & Martínez-Molina, M. (2014). Predictive model of the importance and usefulness of Physical Education. *Cuadernos de Psicología del Deporte*, 14(2), 121-130.
- Baghaei, P., Dourakhshan, A., & Salavati, O. (2012). The relationship between willingness to communicate and success in learning English as a foreign language. *Elixir Psychology*, 53, 12160-12164.
- Baker, S. C., y MacIntyre, P. D. (2000). The role of gender and immersion in communication and second language orientations. *Language Learning*, 50(2), 311-341. <https://doi.org/10.1111/0023-8333.00119>
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of cross-cultural psychology*, 1(3), 185-216. <https://doi.org/10.1177/135910457000100301>
- Brislin, R. W. (1986). Research instruments. *Field methods in cross-cultural research: Cross-cultural research and methodology series*, 8, 137-164.
- Calabuig, F. & Crespo, J. (2009). Uso del método Delphi para la elaboración de una medida de la calidad percibida de los espectadores de eventos deportivos. *Retos: nuevas tendencias en educación física, deporte y recreación*, (15), 21-25.
- Cao, Y., y Philp, J. (2006). Interactional context and willingness to communicate: A comparison of behavior in whole class, group and dyadic interaction. *System*, 34(4), 480-493. <https://doi.org/10.1016/j.system.2006.05.002>
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16-29. <https://doi.org/10.1037/1082-989X.1.1.16>
- Díaz Pinto, E. R. (2009). *Estudio sobre las inteligencias inter-e intrapersonales como instrumentos de desarrollo de la disposición a comunicarse en el aula de inglés*. (Tesis Doctoral), Universidad de Huelva, Huelva.
- Donovan, L. A. & MacIntyre, P. D. (2004). Age and sex differences in willingness to communicate, communication apprehension, and self-perceived competence. *Communication Research Reports*, 21(4), 420-427. <https://doi.org/10.1080/08824090409360006>
- Dörnyei, Z. (2003). *Questionnaires in Second Language Research Construction, Administration, and Processing* (2007 ed.). New Jersey: Lawrence Erlbaum Associates.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis*. New York: Pearson Prentice Hall.
- Hambleton, R., Merenda, P., & Spielberger, C. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. *Adapting educational and psychological tests for cross-cultural assessment*, 1, 3-38.
- Jöreskog, K. G. & Sörbom, D. (2003). LISREL 8.54. *Structural equation modeling with the SIMPLIS command language*. Scientific Software International.
- Lin, Y. & Ranger, A. S. (2003). Sex differences in intercultural communication apprehension, ethnocentrism, and intercultural willingness to communicate. *Psychological Reports*, 92(1), 195-200. <https://doi.org/10.2466/pr0.2003.92.1.195>

- Kamprasertwong, M. (2010). *Willingness to communicate in English speech as a second language: A study of Thai, Chinese, and Dutch samples*. (Trabajo de Fin de Máster. No publicado), University of Groningen, Groningen.
- Kang, S.-J. (2005). Dynamic emergence of situational willingness to communicate in a second language. *System*, 33(2), 277-292. <https://doi.org/10.1016/j.system.2004.10.004>
- MacIntyre, P. D. & Charos, C. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology*, 15(1), 3-26. <https://doi.org/10.1177/0261927X960151001>
- MacIntyre, P. D., Dörnyei, Z., Clément, R., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545-562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- MacIntyre, P. D., Baker, S. C., Clément, R., & Conrod, S. (2001). Willingness to communicate, social support, and language-learning orientations of immersion students. *Studies in Second Language Acquisition*, 23(3), 369-388.
- MacIntyre, P. D., Baker, S. C., Clément, R., & Donovan, L. A. (2002). Sex and age effects on willingness to communicate, anxiety, perceived competence, and L2 motivation among junior high school French immersion students. *Language Learning*, 52(3), 537-564. [10.1111/1467-9922.00226](https://doi.org/10.1111/1467-9922.00226)
- MacIntyre, P. D., & Doucette, J. (2010). Willingness to communicate and action control. *System*, 38(2), 161-171. <https://doi.org/10.1016/j.system.2009.12.013>
- McCroskey, J. C. (1992). Reliability and validity of the willingness to communicate scale. *Communication Quarterly*, 40(1), 16-25. <https://doi.org/10.1080/01463379209369817>
- McCroskey, J. C., y Baer, J. E. (1985). *Willingness to communicate: The construct and its measurement*. Paper presented at the Annual Meeting of the Speech Communication Association, Denver, CO.
- Menezes, E., & Juan-Garau, M. (2015). English learners' willingness to communicate and achievement in CLIL and formal instruction contexts. In M. Juan Garau y J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 221-236). Springer.
- Muñiz, J. (2000). Adaptación de los tests de unas culturas a otras. *Metodología de las Ciencias del Comportamiento*, 2(2), 129-149.
- Muñiz, J. & Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, 12(3), 206.
- Nunnally, J. C. & Bernstein, I. J. (1995). *Teoría psicométrica*. McGraw- Hill.
- Nuviala Nuviala, A., Tamayo Fajardo, J., Iranzo Llopis, J., & Falcón Miguel, D. (2008). Creación, diseño, validación y puesta en práctica de un instrumento de medición de la satisfacción de usuarios de organizaciones que prestan servicios deportivos. *Retos: nuevas tendencias en educación física, deporte y recreación*. (14), 10-16. <https://doi.org/10.47197/retos.v0i14.35004>
- Osterlind, S. J. (1989). *Constructing Test Items*. Kluwer Academic Publishers.
- Peng, J. E. & Woodrow, L. (2010). Willingness to communicate in English: A model in the Chinese EFL classroom context. *Language Learning*, 60(4), 834-876. <https://doi.org/10.1111/j.1467-9922.2010.00576.x>
- Pihko, M. (2007). Foreign language anxiety in content and language integrated learning (CLIL) and in traditional foreign language classes. In S. Tella (Ed.), *From brawn to brain: Strong signals in foreign language education* (pp. 129-142). University of Helsinki.
- Ruiz-Juan, F. & Baena-Extremera, A. (2015). Prediction of achievement goals in Physical Education from the Satisfaction, Motivation and Beliefs of Success in Sport. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 10(2), 193-203.
- Santos Menezes, E. (2014). *Ansiedad y disposición a comunicarse en el aprendizaje del inglés como segunda lengua: estudio de las influencias del modelo formativo (AICLE y Enseñanza Formal)*. (Tesis Doctoral), Universitat de les Illes Balears, Palma.
- Spaan, M. (2006). Test and item specifications development. *Language Assessment Quarterly: An International Journal*, 3(1), 71-79.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum. https://doi.org/10.1207/s15434311laq0301_5

- Vanderveen, T. N. (2015). *Life After French Immersion: Expectations, Motivations, and Outcomes of Secondary School French Immersion Programs in the Greater Toronto Area*. (Tesis Doctoral), York University, Toronto, Toronto, Ontario.
- Wu, H.-p. (2014). How Early Bilingual Education Predicts Mandarin-Chinese Speaking English Language Learners' L2 Oral Communicative Competence Hsiao-ping Wu. *Texas Papers in Foreign Language Education (TPFLE)*, 16(1), 75-88.
- Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and effect on willingness to communicate and second language communication. *Language Learning*, 54(1), 119-152. <https://doi.org/10.1111/j.1467-9922.2004.00250.x>