

Medición de la motivación para aprender inglés como lengua extranjera: Validación del LLOS-IEA. Diferencias de Sexo y Edad en las regulaciones emocionales

Eliseo FERNÁNDEZ-BARRIONUEVO
Javier VILLORIA PRIETO
Francisco Tomás GONZÁLEZ FERNÁNDEZ
María del Mar ORTIZ-CAMACHO
Antonio BAENA-EXTREMERA

Datos de contacto:

Eliseo Fernández-Barrionuevo
Consejería de Educación en la
Embajada de España en EE.UU.
Consulado General de España en
Houston.

eliseo.fernandezb@educacion.gob.es

Javier Villoria Prieto
Universidad de Granada.
Departamento de Didáctica de
la Lengua y la Literatura.
villoria@ugr.es

Francisco Tomás González-
Fernández
Universidad Pontificia de
Comillas. Departamento de
Ciencias de la Actividad Física y
del Deporte
ftgonzalez@comillas.edu

María del Mar Ortiz-Camacho
Universidad de Granada.
Departamento de Didáctica de
la Expresión Musical, Plástica y
Corporal.
mmortiz@ugr.es

Antonio Baena-Extremera
Universidad de Granada.
Departamento de Didáctica de
la Expresión Musical, Plástica y
Corporal.
abaenaextrem@ugr.es

Recibido: 03/07/2021
Aceptado: 26/10/2021

RESUMEN

El objetivo de este estudio ha sido traducir al español, adaptar a un contexto AICLE y analizar las propiedades psicométricas de la escala *LLOS-IEA*; Escala de la orientación del aprendizaje de idiomas – Subescalas de Motivación Intrínseca, Motivación Extrínseca y desmotivación. La versión adaptada del *LLOS-IEA* fue administrada a 3355 estudiantes de Andalucía con edades comprendidas entre los 11 y los 17 años. Tras los procesos de traducción y adaptación, se realizaron análisis de los ítems y de la estructura interna con el fin de comparar 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. También se realizó un análisis de modelos multiniveles para estudiar la validez de constructo obteniendo resultados similares a otros estudios.

PALABRAS CLAVE: Motivación; FLL; LLOS-IEA; validación; sexo; edad.

Measuring motivation to learn in English as a foreign language: LLOS-IEA Validation. Sex and age differences in motivational regulations

ABSTRACT

The object of this research was to translate into Spanish and adapt to a CLIL context the Language Learning Orientation Scale-Intrinsic Motivation, Extrinsic Motivation and Amotivation Subscales (LLOS-IEA) analysing its psychometric properties. The LLOS-IEA 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 comparing the 5 and 7-factor models. This led to a CFA and to reliability and validity analyses, which concluded that the 5-factor performed the best fit. 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; LLOS-IEA; validation; sex; age.

Introducción

Due to the implication in the educational field, motivation in foreign language learning (FLL) has been analyzed from different approaches. The Self-Determination Theory (SDT) is playing a central role in this study as “over the years the theory —and particularly its two linchpins, intrinsic and extrinsic motivation— has become one of the most influential constructs in motivational psychology” (Dörnyei & Ryan, 2015, p. 82).

The SDT is considered a meta-theory of human motivation that studies the degree in which human behaviour is self-determined, it “focuses on types, rather than just amount, of motivation” (Deci & Ryan, 2008, p. 182). This theory sets two main sorts of motivation: intrinsic (IM), which is related with inherent satisfaction of taking part in one activity, and extrinsic (EM), which “refers to the performance of an activity in order to attain some separable outcome” (Ryan & Deci, 2000, p. 71). Both IM and EM have been divided into different sub-categories of motivation (Vallerand, 1997; Vallerand et al., 1989; Vallerand et al., 1992). On the other hand, amotivation (AM) is found when the individuals cannot find any relation between their actions and the consequences of their behaviour. Several studies in the foreign language (FL) field have positively linked IM with some outcomes, such as self-confidence (Pae, 2008), school performance (Morales Rodríguez, 2011), intention to continue with the FL studies and anxiety reduction (Noels et al., 1999; Noels et al., 2000).

This study is based on a Content and Language Integrated Learning (CLIL) setting defined as a “dual-focused educational approach in which an additional language is used for the learning and teaching for both content and language” (Coyle et al., 2010, p. 1). There is much research drawn from different contexts that prove the positive effect of this approach on the student motivation, some examples are the studies of Lasagabaster (2011) in the Basque County, Seikkula-Leino (2007) in Finland, Sylvén and Thompson (2015) in Sweden, Mearns (2012) and Hunt (2011) in the UK or Hewit and García-Sánchez (2012) in the Spanish province of Almeria. However, there is a lack of research based on the analysis of psychometric properties of motivational instruments in Spanish of either FLL and CLIL domains.

Regarding the SDT questionnaire design, the introduction of the “Language Learning Orientation Scale-Intrinsic Motivation, Extrinsic Motivation and Amotivation Subscales” (LLOS-IEA; Noels et al. 2000), opened the door to the study of the IM and EM based on the domain of L2 (second language) learning. This scale was founded on the “Academic Motivation Scale”, (Vallerand et al., 1989; Vallerand et al., 1992). From this point, several scholars designed or/and adapted questionnaires, which proceed from the SDT in the field of L2. Maherzi (2011) proposed a trans-cultural validation for University students in Saudi Arabia. Lucas et al. (2010) created one version with factors for IM regarding the four linguistic skills (speaking, listening, reading and writing). Finally, Ardasheva et al. (2012) designed the English Language Learner Motivation Scale (ELLMS) conceived for pre-university bilingual contexts.

It is necessary to take into account that, during the last decade, the CLIL approach has spread all over Europe, to the extent that most countries have applied it in some of their schools (Eurydice, 2008). In this scenario Coyle (2010) assured that Spain was “rapidly becoming one of the European leaders in CLIL practice and research” (p.xiii), and precisely in the region of Andalusia, this approach has been widely implemented with a total of 947 bilingual schools (Consejería de Educación, 2015a) that served 321.685 CLIL students in the school year 2015 (Consejería de Educación, 2015b). Also, taking into consideration the importance of the Spanish language all over the world, the study of the psychometric properties and the validation of an adapted Spanish LLOS-IEA would open multiple researching opportunities apart from providing an instrument to measure motivation in FL students.

Metod

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).

Instruments

A translation and adaption of the scale designed by Noels et al. (2000) has been implemented. It was based on the “Academic Motivation Scale”, developed by Vallerand et al. (1989), and originally adapted and translated into English by Vallerand et al. (1992). The original instrument assesses the “Amotivation” ($\alpha=.82$), the three types of EM — “External Regulation” ($\alpha=.75$), “Introjected Regulation” ($\alpha=.67$) and “Identified Regulation” ($\alpha=.87$)—, and the three types of IM — “Knowledge” ($\alpha=.85$), “Mastery” ($\alpha=.88$) and “Stimulation” ($\alpha=.85$)—. This scale includes three items for each of its seven factors, where the respondents indicate the extent to which they agree with each item by using a 7-point Likert scale ranged from 1 = Does not correspond at all to 7 = Corresponds exactly.

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 to adapt 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, 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.

The LLOS-IEA was adapted from a FLL context to a context of learning “through” a FL by rephrasing some items when required (e. g. study a second language was replaced by study “in” English). Subsequently, a qualitative evaluation (content validity) of the work was undertaken by five experts (Osterlind, 1989): 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. 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 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=.56 for item suitability and ICC=.83 for item weightiness.

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 (Noels et al., 2000). Subsequently, a second PCA analysis was conducted with no fixed factors to extract.

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

LLOS-IEA 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, West, and Finch (1996) (table 1).

Table 1

Statistical analysis of each item of LLOS-IEA

Scale: LLOS-IEA (n = 3355)	M	SD	CITC-c	CC	α without item	Skewness	Kurtosis
Amotivation ($\alpha=.783$)							
4. No llego a entender por qué estudio en inglés y sinceramente me da igual.	1.75	1.48	.64	.85	.68	2.16	3.80
10. No lo sé. No puedo llegar a entender por qué tengo que estudiar en inglés	1.74	1.50	.65	.86	.67	2.20	3.92
19. Honestamente no lo sé, tengo la impresión de estar perdiendo el tiempo cuando estudio en inglés.	1.58	1.33	.58	.80	.75	2.58	6.00
EM-external regulation ($\alpha=.787$)							
14. Para conseguir un mejor trabajo en el futuro.	6.43	1,09	.62	.90	.73	-2.44	6.53
20. Para conseguir un trabajo más prestigioso en el futuro.	6.26	1.24	.73	.93	.61	-2.09	4.41
21. Para tener un mejor sueldo en el futuro.	5.89	1,55	.58	.58	.81	-1.52	1.64
EM-intojected regulation ($\alpha= .694$)							
9. Porque me sentiría culpable si no consiguiera hablar en inglés.	3.34	2.12	.56	.89	.53	.38	-1.24
11. Porque me daría vergüenza si no pudiera hablar en inglés.	3.10	2.02	.55	.87	.55	.57	-.95
15. Porque me siento importante cuando puedo hablar en inglés.	4.36	1.96	.42	.42	.71	-.27	-1.07
EM-identified regulation ($\alpha=.631$)							

1. Porque creo que es bueno para mi desarrollo personal.	5.80	1.38	.33	.33	.80	-1.13	.90
5. Porque elijo ser la clase de persona que puede hablar más de una lengua.	5.34	1.76	.59	.92	.53	-.90	-.17
6. Porque elijo ser el tipo de persona que puede hablar inglés.	5.43	1.64	.61	.91	.41	-.93	.03
IM-knowledge ($\alpha=.692$)							
2. Por el sentimiento de satisfacción que tengo al descubrir nuevas cosas.	4.91	1.60	.49	.75	.62	-.47	-.45
16. Por el placer que siento sabiendo más acerca de la música, series o literatura en inglés.	4.70	1.94	.48	.80	.64	-.45	-.95
18. Porque me gusta la sensación de adquirir conocimientos sobre lugares donde se habla inglés y sobre su forma de vida.	4.94	1.78	.56	.82	.53	-.55	-.70
IM-stimulation ($\alpha=.840$)							
3. Por la sensación que experimento cuando hablo en inglés.	4.53	1.75	.65	.84	.82	-.31	-.77
7. Por la sensación que siento cuando oigo hablar en inglés.	4.29	1.77	.76	.89	.73	-2.17	-.88
8. Por el placer que siento al escuchar el inglés por hablantes nativos.	4.22	1.92	.71	.88	.78	-.15	-1.09
IM-accomplishment ($\alpha=.803$)							
12. Por el placer que siento cuando me supero a mí mismo en los estudios en inglés	5.43	1.67	.68	.86	.70	-.93	.01
13. Por la satisfacción que siento cuando estoy resolviendo ejercicios difíciles en inglés.	5.02	1.86	.67	.87	.71	-.70	-.54
17. Por el placer que experimento cuando comprendo un concepto diferente sobre algo en inglés.	4.93	1.71	.60	.85	.79	-.57	-.48

Items from factor 1 (AM) showed M values between 1.58 from item 19 and 1.75 from item 4. All the SD values were >1 , and this dimension's internal consistency was satisfactory ($\alpha=.783$). All the CITC-c were $\geq .58$.

With regard to EM, all the items from factor 2 (EM-external regulation) presented mean values from 5.89 (item 21) to 6.43 (item 14), SD values were >1 , and this dimension's internal consistency was suitable ($\alpha=.787$). However, the removal of item 21 would increase α up to $.81$. All the CITC-c were $\geq .58$. Relative to factor 3 (EM-introjected regulation) all the mean values ranged between 3.10 (item 11) to 4.36 (item 15). SD values were >1 , and this dimension's internal consistency was nearly satisfactory ($\alpha=.694$), but the dismissal of item 15 would raise α up to $.71$. All the CITC-c values were $\geq .42$. Finally, factor 4 (EM-identified regulation) showed values from 5.34

(item 5) to 5.80 (item 1). SD values were >1 and this dimension's internal consistency was nearly satisfactory ($\alpha=.631$) nevertheless the α value of this factor without item 1 would be .80. All the CITC-c were $\geq .33$.

Concerning the IM, all the items from factor 5 (IM knowledge) ranged from 4.70 (item 16) to 4.94 (item 18). The SD values were >1 , and this dimension's internal consistency was nearly satisfactory ($\alpha=.692$). All the CITC-c values were $\geq .48$. On the other hand, factor 6 (IM-stimulation) mean values fluctuated between 4.22 (item 8) and 4.53 (item 3). The internal consistency was satisfactory ($\alpha=.840$), and all the CITC-c values were $\geq .65$. Finally, in factor 7 (IM-accomplishment) all the mean values were between 4.93 (item 17) and 5.43 (item 12). The internal consistency was satisfactory ($\alpha=.803$), and all CITC-c values were $\geq .60$.

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 $CC \geq .33$. It is worth mentioning the fact that items 4, 10, 19, 14 y 20 presented higher Kurtosis or Skewness values which were taken into consideration for the studies mentioned later on. However, Curran et al. (1996) state that the Kurtosis value could reach as far as 7.0. The rest of the items are ranged within the acceptable values so as for avoiding their removal.

Internal structure analysis

Following the validation process of the original instrument (Noels et al., 2000), a EFA for the seven-factor model was first 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 (.892), and the Bartlett's sphericity test was significant ($c^2_{(210)}=29186.601$, $p<.000$), concluding with the suitability of the implementation of the EFA. The results confirmed a 7-factors extraction accounting for 70.432% of the total variance explained (table 2).

The results confirmed a dimensional 7-factor structure with saturation values above .41. However, items from IM-knowledge and IM-achievement merged into one dimension. On the other hand, contrary to the original instrument, items 1 and 2 would be placed respectively outside the EM-identified regulation and IM-knowledge dimensions. It is worth mentioning that the item 2 factor saturation value was .41.

Bearing in mind that both merged factors belong to the IM group, and that some other investigations have analysed 5-factors models for similar instruments (Li & Harmer, 1996; Ntoumanis, 2001), a new PCA analysis was conducted. In this case, no fixed factors to extract were indicated, requiring at least a .40 correlation in order to consider each item important within the factor (Stevens, 1992). The KMO index was good (.890) and Bartlett's sphericity test was significant ($c^2_{(210)}=29278.425$ $p<.000$), concluding with the suitability of the implementation of the EFA. The results of this analysis yielded seven factors, accounting for 62.287% of the total variance explained (table 2). It should be noted that item 1, which was formerly situated outside the

theoretical model, now belongs to IM factor, although presenting lower h^2 and factor saturation values—a fact to be considered in future analysis.

Table 2

LLOS-IEA 5 and 7-model factor saturation and communalities

Item:	7 factor extraction			5 factor extraction		
	S	h^2	D	S	h^2	D
1. Porque creo que es bueno para mi desarrollo personal.	.82	.80	X	.46	.34	IM
2. Por el sentimiento de satisfacción que tengo al descubrir nuevas cosas.	.60	.67	X	.73	.56	IM
3. Por la sensación que experimento cuando hablo en inglés.	.72	.71	MI-ES	.75	.61	IM
4. No llego a entender por qué estudio en inglés y sinceramente me da igual.	.82	.72	AM	.81	.70	AM
5. Porque elijo ser la clase de persona que puede hablar más de una lengua.	.86	.85	EM-ID	.84	.82	EM-ID
6. Porque elijo ser el tipo de persona que puede hablar inglés.	.80	.82	EM-ID	.78	.81	EM-ID
7. Por la sensación que siento cuando oigo hablar en inglés.	.80	.80	IM-ST	.73	.63	IM
8. Por el placer que siento al escuchar el inglés por hablantes nativos.	.76	.74	IM-ST	.72	.60	IM
9. Porque me sentiría culpable si no consiguiera hablar en inglés.	.82	.72	EM-IN	.81	.69	EM-IN
10. No lo sé. No puedo llegar a entender por qué tengo que estudiar en inglés	.84	.73	AM	.84	.73	AM
11. Porque me daría vergüenza si no pudiera hablar en inglés.	.85	.75	EM-IN	.83	.70	EM-IN
12. Por el placer que siento cuando me supero a mí mismo en los estudios en inglés	.72	.66	IM-AC/ IM-KN	.64	.54	IM
13. Por la satisfacción que siento cuando estoy resolviendo ejercicios difíciles en inglés.	.69	.60	IM-LO/ IM-CO	.65	.53	MI

14. Para conseguir en un mejor trabajo en el futuro.	.75	.71	EM-EX	.78	.71	EM-EX
15. Porque me siento importante cuando puedo hablar en inglés.	.54	.53	EM-IN	.58	.49	EM-IN
16. Por el placer que siento sabiendo más acerca de la música, series o literatura en inglés.	.66	.53	IM-AC/ IM-KN	.59	.41	IM
17. Por el placer que experimento cuando comprendo un concepto diferente sobre algo en inglés.	.76	.70	IM-AC/ IM-KN	.73	.61	IM
18. Porque me gusta la sensación de adquirir conocimientos sobre lugares donde se habla inglés y sobre su forma de vida.	.64	.60	IM-AC/ IM-KN	.70	.53	IM
19. Honestamente no lo sé, tengo la impresión de estar perdiendo el tiempo cuando estudio en inglés.	.78	.66	AM	.78	.64	AM
20. Para conseguir un trabajo más prestigioso en el futuro.	.86	.80	EM-EX	.86	.79	EM-EX
21. Para tener un mejor sueldo en el futuro.	.84	.73	EM-EX	.79	.66	EM-EX

S=factor saturation, h²=communalities, D=dimension, X=out-of-theoretical-model-component, IM=intrinsic motivation, IM-AC=intrinsic motivation-accomplishment, IM-KN= intrinsic motivation-knowledge, IM-ST=intrinsic motivation-stimulation, EM-EX=extrinsic motivation-external regulation, EM-IN=extrinsic motivation-introyected regulation, EM-ID=extrinsic motivation-identified regulation, AM=amotivation.

Confirmatory factor analysis

In order to study the psychometric properties of LLOS-IEA original dimesionalization (Noels et al., 2000), structural equation modelling was performed. Different absolute and relative fitness indices were calculated (Bentler, 2007; Markland, 2007), such as p-value associated with Chi-square test, c² and degrees of freedom ratio (df; c² /df), goodness of fit index (GFI), normed fit index (NFI), non-normed fix 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 .317. 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 and Sörbom, 2003) program was conducted. The polychoric correlations matrix and asymptotic covariance matrix were used as input for data analysis. Following Markland's (2007) suggestions of formulating several models if recommended by data (see tables 1 and 2), three different models were hypothesized. Model 1 would follow the item-factor distribution from the original

instrument; model 2 would dismiss the items whose removal would improve the α value (items 1, 15, 17 and 21); and model 3 would consider items 2, 11, 16, 17, 18, 14, 20, 4, 10, 19, 9, 11, 5 and 6. The former was a five-factor model (figure 1) in line with other works in the sport field (Li & Harmer, 1996; Ntoumanis, 2001) or FLL domain, such as Ardasheva et al. (2012) instrument, which contained a unique factor for IM.

Most of the items presented individual reliability (R^2) values $>.5$ in the three models, being the lowest value in item 16 for the model 1 ($R^2=.431$), and in item 2 for the model 2 ($R^2=.431$).

Table 3 displays the three model goodness of fit. Nevertheless, only in the five-factor model the χ^2/df value was <5.00 , confirming this model fit (Hu & Bentler, 1999). Attending each model cross-validation index (ECVI) and the Akaike information criterion—which apart from the fit, informs about the model parsimony—the five-factor model performed the best fit as its values were the lowest.

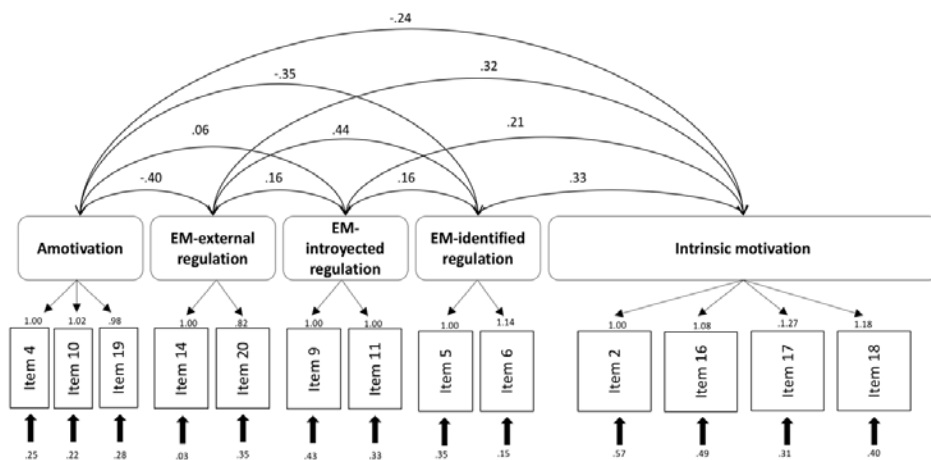
Table 3

Fit indices for each of the LLOS-IEA models

	χ^2	df	χ^2/df	p	GFI	NFI	NNFI	CFI	RMSEA	ECVI	Akaike
Model 1	1851.3	168	11.01	<.000	.98	.93	.93	.95	.05	9.48	671.271
Model 2	561.27	98	5.72	<.000	.99	.98	.98	.98	.03	7.76	439.786
Model 3	205.40	54	4.63	<.000	.99	.99	.99	.99	.03	5.06	279.403

Figure 1

Path diagram of the CFA, with standardized weights and measurement errors of each one of the items in the Spanish LLOS-IEA version for CLIL students.



Reliability and validity

Table 4 shows the model 3 reliability and validity. Regarding model 2, expect for the χ^2 df data, it also performed a good fit therefore its results are also presented as they might be taken into consideration for future studies.

Factors 3 and 4 in model 3 could be questioned due to its $\alpha < .70$, nevertheless, as a consequence of the limited items per factor (2), these values would be acceptable (Taylor et al., 2008). Moreover, the composite reliability is considered more suitable than the Cronbach's alpha, as it does not depend on the number of attributes associated to each concept (Vandenberg, 1996); therefore, factors 3 and 4 in model 3 would show a positive reliability.

Table 4

Reliability and validity of the scale

Dimensions	7-factor model			5-factor model		
	Composite reliability	AVE	α	Composite reliability	AVE	α
Amotivation	.90	.75	.78	.90	.75	.78
EM-external regulation	.90	.82	.79	.89	.81	.79
EM-introyected regulation	.77	.63	.69	.80	.66	.69
EM-identified regulation	.87	.76	.63	.86	.75	.63
IM-knowledge	.77	.53	.69			
IM-stimulation	.89	.73	.84			
IM-accomplishment	.85	.74	.80			
MI				.82	.56	.89

Convergent validity

Regarding convergent validity, some authors such as Bollen (1989) assure that the index validity can be estimated from the magnitude of factor loadings in addition to the adjusted goodness of fit index (AGFI; table 5). In addition to this argument, as stated before, all the factor saturations were statistically significant (t-value >1.96), hence, all the indices assessed the same theoretical construct. Finally, it is worth mentioning that the factor loading for all the items in model 3 were high ($R^2 > .50$).

Table 5

Convergent validity of the scale

	NFI	AGFI
Model 1	.94	.98
Model 2	.98	.98
Model 3	.99	.99

Construct validity

At last, a multi-level model analysis was performed. 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 6 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 to the statistical tests of margin corrected means comparison by multiple comparisons through SIDAK.

Table 6

Mixed model multi-level analysis outcome

		Adjusted mean	Standard Error	95% Confidence interval		Adjusted differences with the reference value	multilevel analysis statistical tests			SIDAK
				Lower	Upper		F	df	p-value	p-value
AM	Boy	1.940	.067	1.808	2.072	.310*	56.606	3351	.000	.000
	Girl	1.630	.067	1.499	1.760					
	1 st cycle	1.723	.026	1.673	1.774	.079	2.878	3351	.056	.190

	2 nd cycle	1.644	.035	1.576	1.713	-.343				.197
	Bachillerato	1.987	.186	1.622	2.352	.263				.409
EM- external regulation	Boy	6.146	.06	6.028	6.264	-.153*	17.151	3351	.000	.000
	Girl	6.299	.059	6.182	6.416					
	1 st cycle	6.400	.023	6.355	6.445	.151*	9.554	3351	.000	.000
	2 nd cycle	6.249	.031	6.188	6.310	.230				.436
	Bachillerato	6.019	.166	5.693	6.345	-.381				.068
EM- introjected regulation	Boy	3.371	.103	3.169	3.573	.016	.067	3351	.796	.796
	Girl	3.355	.102	3.155	3.554					
	1 st cycle	3.184	.039	3.107	3.262	-.085	1.906	3351	.149	.484
	2 nd cycle	3.270	.053	3.165	3.374	-.365				.502
	Bachillerato	3.635	.285	3.076	4.193	.451				.312
EM- identified regulation	Boy	5.046	.088	4.873	5.219	-.113*	4.376	3351	.037	.037
	Girl	5.159	.087	4.988	5.329					
	1 st cycle	5.423	.034	5.357	5.489	.083	7.108	3351	.001	.366
	2 nd cycle	5.339	.045	5.250	5.429	.795*				.004
	Bachillerato	4.545	.243	4.068	5.021	-.878*				.001
IM	Boy	4.565	.076	4.415	4.715	-.379*	65.73	3351	.000	.000
	Girl	4.944	.075	4.796	5.092					
	1 st cycle	4.943	.029	4.885	5.00	.229*	11.640	3351	.000	.000
	2 nd cycle	4.713	.039	4.636	4.791	.105				.947
	Bachillerato	4.608	.211	4.194	5.022	-.335				.310

Concerning the AM dimension, the only differences were observed in the student sex ($p < .000$), being higher in boys ($M = 1.940$; $SE = .067$, adjusted difference = $.310$) with a very important F value ($F = 56.60$).

With regard to the EM-external regulation, there were differences by sex ($p < .000$) and age ($p < .000$) being respectively higher in girls and 1st cycle students.

Concerning the EM-identified regulation there were only differences by age, specifically in 2nd cycle pupils ($M = 5.339$, $SE = .045$), being $.795$ higher than bachillerato students ($M = 4.545$, $SE = .243$).

In the variable EM-introjected regulation, no significant differences were found. Finally, important differences were found by sex ($F = 65.73$; $p < .000$), and age ($F = 65.73$; $p < .000$) for IM, being the higher mean value in first cycle students ($M = 4.943$; $SE = .029$).

Sex Invariance

In order to analyze the factorial invariance, Abalo et al., (2006) recommendations were followed estimating the same model for both samples. No significant differences were found in χ^2 between models, rejecting the H_0 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 $\Delta CFI \leq .01$ 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.

Discussion

The main objective in this work has been to study and validate the LLOS-IEA to a Spanish speaking context. Taking the study of Fernández-Barrionuevo and Baena-Extremera, 2018 as a starting point, this instrument validation will allow us to obtain information about the student FL motivation in CLIL and regular schools. This fact is of great interest due to the expansion of the Spanish language with a total of about 560 million speakers all over the world.

After a first internal structure analysis, a 7 factor-dimensional structure similar to the original instrument (Noels et al., 2000) was to account for 70.432% of the variance. Nevertheless, several factors presented some issues worth considering. First, the IM-accomplishment and IM-knowledge merged in a unique dimension, and second, items 1 and 2 were grouped in one only dimension situated outside the theoretical model. Because of this, a second internal structure analysis was performed indicating no fixed factors to extract, yielding 5 dimensions, in which all the items were placed in their corresponding dimension accounting for 62.287% of the total variance.

As be observed in table 1, the removal of items 1, 15, 17 and 21 would improve Cronbach's alpha values in their respective factors. This was taken into consideration by conducting a 7-factor CFA without these items (model 2) compared to the structure of the original 7-factor instrument (model 1) and a 5-factor version (model 3). Regarding these proposals, the lowest ECVI y Akaike values were in model 3, which indicated the best fitting. Moreover, even though the model 2 χ^2 df value was close to 5,

only the model 3 presented—as recommended—a value under 5.

Although composite reliability and AVE values are above the minimum required in models 2 and 3 (Hair et al., 2009), the former not only shows the best fit, but also better convergent validity. However, model 2 should also be considered for future research as its CFA and its validity and reliability indices are close to acceptable.

Even though we have not found any study based on a LLOS-IEA 5-factor model, there are some other instruments also found on the EME (Vallerand et al., 1989) with this factor structure such as the SMS Spanish 5-factor version (Duda, 2007; Granero-Gallegos et al., 2014). Moreover, as regards the FL, the scale designed by Ardasheva et al. (2012) shares some features with our instrument, for instance, the use of a unique factor for IM and the contextualization in pre-university bilingual students.

In relation to the construct validity, except for the EM-introjected regulation, all the motivational variables were significantly higher for girls, while AM was significantly higher for boys. These results are in line with other studies that also indicate higher IM values in schoolgirls (Kissau, 2006; Okuniewski, 2014; Williams et al., 2002). The cause of these sex differences could be in the feminized bias of the FL school domain (Williams et al., 2002). This idea is backed up by the McCall's (2011) study where the combination of FLL with one interesting topic for boys such as football, improved their IM.

Regarding the higher FL motivation in girls, Williams et al. (2002) stated that “comments elicited within individual interviews suggest that this [the positive attitude towards FLL] might be part of a more general orientation toward schoolwork rather than necessary relating only to learning languages” (p. 522). Therefore, another cause of the higher FL motivation in girls could be caused by a higher academic motivation in general. Based on Vallerand (1997) Hierarchical Model, motivation in a certain context, such as the school, could have an effect on another context, for instance the FLL.

To conclude, results indicate that the 5-factor Spanish version of LLOS-IEA adapted for CLIL schools is valid, reliable and a sex-invariant instrument. However, it would be advisable to test the 7-factor model in future investigations. This instrument will open new fields of research by allowing us to obtain motivational information in CLIL and FLL Spanish speaking students that could not be accessed before. In this way, investigations will be able to deepen in one of the most important FLL underlying factors—student's motivation, facing the study “through” a FL as a specific domain. In addition, teachers will have a valid tool to measure student's motivation when learning a/through a FL.

Conflict of interests

Los autores declaran no tener ningún conflicto de intereses. Los financiadores no tuvieron ningún papel en el diseño del estudio; en la recopilación, análisis o interpretación de datos; en la redacción del manuscrito, o en la decisión de publicar los resultados.

Authors' contributions

Los artículos de investigación elaborados por varios autores deben especificar brevemente sus contribuciones individuales para las que pueden utilizarse las siguientes declaraciones: Conceptualización, E.F.B., J.V.P. y A.B.E metodología, E.F.B y J.V.P.; validación, E.F.B., J.V.P. y

A.B.E.; análisis formal, E.F.B. y A.B.E.; análisis de datos, M.M.O.C.; redacción del borrador original, E.F.B., J.V.P. y A.B.E.; redacción, revisión y edición, F.T.G.F.; supervisión, J.V.P. y A.B.E.

References

- Abalo, J., Lévy, J., Rial, A., y Varela, J., *Invarianza factorial con muestras múltiples*. En J. Lévy (Ed.), *Modelización con Estructuras de Covarianzas en Ciencias Sociales* (pp. 259-278). Netbiblo, Madrid (2006).
- Ardasheva, Y., Tong, S. S., & Tretter, T. R. (2012). Validating the English Language Learner Motivation Scale (ELLMS): Pre-college to measure language learning motivational orientations among young ELLs. *Learning and Individual Differences, 22*(4), 473-483. doi: 10.1016/j.lindif.2012.03.001
- Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42*(5), 825-829. doi: 10.1016/j.paid.2006.09.024
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of cross-cultural psychology, 1*(3), 185-216. doi: 10.1177/135910457000100301
- Brislin, R. W. (1986). Research instruments. *Field methods in cross-cultural research: Cross-cultural research and methodology series, 8*, 137-164. doi: 10.2307/2349027
- 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, 0*(15), 21-25. Retrieved from <https://recyt.fecyt.es/index.php/retos/article/view/34993>
- Carretero, H., & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology, 5*(3), 521-551.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. doi: 10.1207/S15328007SEM0902_5
- Consejería de Educación (2015a). Portal Averroes. Buscador de proyectos en los centros docentes. Retrieved from <http://www.juntadeandalucia.es/educacion/portalaverroes>
- Consejería de Educación (2015b). *La Educación en Andalucía. Avance curso 2015-2016. Iniciativas, programas y datos*. Sevilla: Consejería de Educación. Junta de Andalucía.
- Coyle, D. (2010). Foreword. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training* (pp. vii-viii). New Castle, UK: Cambridge Scholars Publishing.
- 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. doi: 10.1037/1082-989X.1.1.16

- Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne*, 49(3), 182. doi: 10.1037/a0012801
- Declaration of Helsinki (2008). World Medical Association [online]. Retrieved from <http://www.wma.net/es/30publications/10policias/b3/>
- Dörnyei, Z. (2003). *Questionnaires in Second Language Research Construction, Administration, and Processing* (2007 ed.). New Jersey: Lawrence Erlbaum Associates.
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*: Routledge.
- Duda, J. L. (2007). Propiedades psicométricas de la escala de motivación deportiva en deportistas españoles. *Revista Mexicana de Psicología*, 24(2), 197-207.
- Eurydice. (2008). *Cifras clave de la enseñanza de lenguas en los centros escolares de Europa*. Instituto de Formación del Profesorado Investigación e Innovación Educativa IFIIE.
- Fernández-Barrionuevo, E., & Baena-Extremera, A. (2018). Motivation in Physical Education and Foreign Language Learning in CLIL Education. Implication for future studies. *Porta Linguarum*. 30: 207-220.
- Granero-Gallegos, A., Baena-Extremera, A., Gómez-López, M., Sánchez-Fuentes, J. A., & Abrales, J. A. (2014). Psychometric Properties of the "Sport Motivation Scale (SMS)" Adapted to Physical Education. *Journal of Sports Science & Medicine*, 13(4), 801-807. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4234949/>
- 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.
- Hewit, E. C., & García-Sánchez, E. (2012). Evolución del Aprendizaje Integrado de Contenidos y Lengua (AICLE/CLIL) en España: Un Proyecto Empírico en la Universidad. *Latin American Journal of Content Language Integrated Learning*, 5(1), 57-67. doi: 10.5294/laclil.2012.5.1.5
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation modeling*, 6, 1-55. doi: 10.1080/10705519909540118
- Hunt, M. (2011). Learners' perceptions of their experiences of learning subject content through a foreign language. *Educational Review*, 63(3), 365-378. doi: 10.1080/00131911.2011.571765
- Kissau, S. (2006). Gender differences in second language motivation: An investigation of micro-and macro-level influences. *Canadian Journal of Applied Linguistics*, 9(1), 73-96. Retrieved from: <https://journals.lib.unb.ca/index.php/CJAL/article/view/19755/21452>
- Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language Learning and Teaching*, 5(1), 3-18. doi: 10.1080/17501229.2010.519030

- Li, F., & Harmer, P. (1996). Testing the simplex assumption underlying the Sport Motivation Scale: A structural equation modeling analysis. *Research Quarterly for Exercise and Sport*, 67(4), 396-405. doi: 10.1080/02701367.1996.10607971
- Lucas, R. I., Pulido, D., Miraflores, E., Ignacio, A., Tacay, M., & Lao, J. (2010). A study on the intrinsic motivation factors in second language learning among selected freshman students. *Philippine ESL Journal*, 4, 3-23. Retrieved from: http://www.philippine-esl-journal.com/wp-content/uploads/2014/01/V4_A1.pdf
- Maherzi, S. (2011). Perception of classroom climate and motivation to study English in Saudi Arabia: Developing a questionnaire to measure perception and motivation. *Electronic Journal of Research in Education Psychology*, 9(2), 763-798. Retrieved from: <http://www.investigacion-psicopedagogica.org/revista/new/english/ContadorArticulo.php?554>
- Markland, D. (2007). The golden rule is that there is no rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences*, 42, 852-858. doi: 10.1016/j.paid.2006.09.023
- McCall, I. (2011). Score in French: motivating boys with football in Key Stage 3. *Language Learning Journal*, 39(1), 5-18.
- Mearns, T. L. (2012). Using CLIL to enhance pupils' experience of learning and raise attainment in German and health education: a teacher research project. *The Language Learning Journal*, 40(2), 175-192. doi: 10.1080/09571736.2011.621212
- Morales Rodríguez, F. M. (2011). Aprendizaje, Motivación y Rendimiento en Estudiantes de Lengua Extranjera Inglesa. *Psicología Educativa*, 17(2), 195-207. doi: 10.5093/ed2011v17n2a6
- 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. doi: 10.1027/1016-9040.12.3.206
- Noels, K. A., Clement, R., & Pelletier, L. G. (1999). Perceptions of teachers' communicative style and students' intrinsic and extrinsic motivation. *Modern Language Journal*, 83(1), 23-34. doi: 10.1111/0026-7902.00003
- Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language learning*, 50(1), 57-85. doi: 10.1111/1467-9922.53223
- Ntoumanis, N. (2001). A self-determination approach to the understanding of motivation in physical education. *British Journal of Educational Psychology*, 71(2), 225-242. doi: 10.1348/000709901158497
- Nunnally, J. C., & Bernstein, I. J. (1995). *Teoría psicométrica*. Madrid: 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. Retrieved from: http://retos.org/numero_14_13/retos%2014-2.pdf

- Okuniewski, J. E. (2014). Age and Gender Effects on Motivation and Attitudes in German Learning: The Polish Context. *Psychology of Language and Communication, 18*(3), 251-262. doi: 10.2478/plc-2014-0017
- Osterlind, S. J. (1989). *Constructing Test Items*. Londres: Kluwer Academic Publishers.
- Pae, T.-I. (2008). Second language orientation and self-determination theory - A structural analysis of the factors affecting second language achievement. *Journal of Language and Social Psychology, 27*(1), 5-27. doi: 10.1177/0261927X07309509
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68. doi: 10.1037/0003-066X.55.1.68
- Seikkula-Leino, J. (2007). CLIL Learning: Achievement Levels and Affective Factors. *Language and Education, 21*(4), 328-341. doi: 10.2167/le635.0
- Spaan, M. (2006). Test and item specifications development. *Language Assessment Quarterly: An International Journal, 3*(1), 71-79. doi: 10.1207/s15434311laq0301_5
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hills- Dale, NJ: Lawrence Erlbaum.
- Sylvén, L. K., & Thompson, A. S. (2015). Language learning motivation and CLIL: Is there a connection? *Journal of Immersion and Content-Based Language Education, 3*(1), 28-50. doi: 10.1075/jicb.3.1.02syl
- Taylor, I., Ntoumanis, N., & Standage, M. (2008). A self-determination theory approach to understanding antecedents of teachers' motivational strategies in physical education. *Journal of Sport and Exercise Psychology, 30*, 75-94. doi: 10.1123/jsep.30.1.75
- Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. *Advances in experimental social psychology, 29*, 271-360. doi: 10.1016/S0065-2601(08)60019-2
- Vallerand, R. J., Blais, M. R., Brière, N. M., & Pelletier, L. G. (1989). Construction et validation de l'échelle de motivation en éducation (EME). *Canadian Journal of Behavioural Science, 21*(3), 323-349. doi: 10.1037/h0079855
- Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C., & Vallières, E. F. (1992). The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement, 52*(4), 1003-1017. doi: 10.1177/0013164492052004025
- Vandenbosch, M. (1996). Confirmatory compositional approaches to the development of product spaces. *European Journal of Marketing, 30*(3), 23-46. doi: 10.1108/03090569610107418
- Williams, M., Burden, R., & Lanvers, U. (2002). 'French is the language of love and stuff': Student perceptions of issues related to motivation in learning a foreign language. *British Educational Research Journal, 28*(4), 503-528. doi: 10.1080/0141192022000005805