Revista ELectrónica de Investigación y EValuación Educativa

RELIEVE

e-Journal of Educational Research, Assessment and Evaluation

ISSN: 1134-4032

Linguistic diversity and assessment in multilingual schools

Diversidad lingüística y evaluación en centros multilingües

Elosua, Paula
Universidad del País Vasco

Abstract

In a context where linguistic diversity in the classroom is natural, educational assessment raises new challenges that demand the conjunction of different perspectives. The equity in the assessment process requires considering facets related to the psychometric properties of the tests, the development of linguistic competence, or the sociolinguistic context through the incorporation of context variables into the predictive models of performance. By analyzing population data from compulsory secondary education schools which offer more than one linguistic educational model in the Autonomous Community of the Basque Country (N = 45), this paper examines the responses of 2783 students to a mathematical competence test. The methodological approach is twofold; it includes a psychometric study of equivalence through the estimation of the differential item functioning with regard to the family language and the linguistic educational model, and the estimation of mixed linear models for the prediction of the mathematics competence. The results show: a) the equivalence of the assessment test in relation to the family language and educational model, b) the significant impact of the linguistic context variables on the competence estimation, c) the lack of significance of the linguistic models on performance. It is concluded the importance of conducting studies of differential functioning, and the incorporation of linguistic variables in the prediction of mathematical competence.

Keywords: educational assessment, student diversity, language of instruction, differential item functioning.

Reception Date
2018 January 25

Approval Date
2018 December 05

Publication Date:
2018 December 05

Resumen

En un entorno en el que la diversidad lingüística es natural, la metodología de evaluación afronta nuevos retos. La equidad en el proceso evaluativo aconseja la consideración de aspectos relacionados con las características psicométricas de las pruebas, el desarrollo de la competencia lingüística, o el contexto sociolingüístico, a través de modelos explicativos sobre las variables contextuales de naturaleza lingüística que inciden en el rendimiento. Utilizando datos poblacionales de centros educativos de segundo curso de enseñanza secundaria obligatoria que ofertan más de un modelo lingüístico en la Comunidad Autónoma del País Vasco (N=45), este trabajo analiza las respuestas de 2783 estudiantes a una prueba de competencia matemática y estudia el efecto de la lengua familiar (euskera o castellano). La aproximación metodológica incluye un doble acercamiento; un estudio psicométrico de equivalencia a través del análisis del funcionamiento diferencial del ítem en función de lengua familiar y del modelo educativo, y el ajuste de modelos lineales mixtos para la predicción de la competencia. Los resultados muestran: a) la equivalencia de la prueba de evaluación en función del idioma familiar y del modelo educativo, b) el impacto significativo de variables de contexto lingüístico, y c) la falta de significatividad del modelo lingüístico sobre el desempeño. Se concluye la importancia de llevar a cabo estudios de funcionamiento diferencial, y de la incorporación de variables de naturaleza lingüística en la predicción de la competencia matemática.

***Palabras clave*:** evaluación educativa, diversidad, lenguaje de instrucción, funcionamiento diferencial del ítem.

Fecha de recepción
2018 Enero 25

Fecha de aprobación
2018 Diciembre 05

Fecha de publicación
2018 Diciembre 05

Linguistic diversity in the classroom is far from being a rarity or an unusual situation; it is a common scene in our educational practice. The revitalization projects of native languages, the inclusion of multilingual programs that incorporate foreign languages in classrooms,

Autor de contacto / Corresponding author

Elosua, Paula. Department of Social Psychology and Methodology in Behavioral Sciences. University of the Basque Country (Spain). paula.elosua@ehu.es

migration, and globalization in general, define diverse educational spaces.

In a country with 17 autonomous regions and five official languages that coexist with Spanish, the classroom is multilingual. Aranese, Catalan, Galician, Basque, and Valencian are their own languages in six autonomous regions, and the right to know them, use them, and receive teaching in them is legally recognized. With the common objective of revitalization of the minority languages, every autonomous community has designed linguistic models to boost the knowledge to a level that develops the students' linguistic competencies. The linguistic models are shot; from the total immersion in Catalonia, for example, until the coexistence of three different models in the Basque country. A legislative review that protects every linguistic model reveals that:

a) Educational law in Catalonia (Law 12/2009 July 10) defines Catalan as medium of instruction

b) In Valencia, the multilingual law (Decree 9/2017 January 27th; although partially cancelled by the Superior Court of Valencia) set three linguistic levels in function with the nature of the teaching language

c) In the Balearic Islands the foreign language decree (Decree 9/2017 January 27th) allows every school to design their own linguistic project and, according to it, propose the proportion of which each language is used as a teaching language

d) In Galicia the decree 79/2010, May 20th, determines which languages are taught in some subjects, giving freedom to the school to choose the language for the rest, always guaranteeing that the subjects in Galician and Spanish are distributed the same number of hours weekly

e) In the Basque country, the decree 138/1983, July 11th, outlines three linguistic models that can coexist in the same school. The teaching language can be Spanish, Basque, or both

f) The Navarran community adapted the Basque system for the Basque areas and added a new model for the areas without a Basque presence (Navarran Law 4/2015 March 10th).

The implementation of native languages in the educational system has had to respond to new challenges. These include among others promoting teacher proficiency in the minority language, developing educational materials in the minority language, and the language standardization. Next to these, the educational assessment also faces specific challenges to those who must respond with adequate methodologies. The attention to diversity requires considering that in a group of students:

a) The first language and the instruction language can differ

b) The level of linguistic competency in the language of assessment can be insufficient

c) The assessment tools can be presented in more than one language and/or

d) The languages in contact can vary in many important characteristics such as their status and prestige.

These aspects in conjunction are a clear reflection of multifaceted nature that characterizes the assessment in contexts of diversity; it covers disciplines such as psycholinguistics, psychometrics, and sociolinguistics (Elosua, 2016). From every one of these it is possible to consider potential sources of bias that interact and threaten the equity of the assessment process, and that they can be tied to the student's linguistic competence, the characteristics of the assessment tool, and the sociolinguistic environment.

The linguistic competence as a source of bias makes reference to the minimal level of competency or threshold (Cummins, 1981, 2000) that a student should reach to function successfully in a second-language classroom. The Cognitive academic language proficiency (CALP) would be reached between five and eight years, and would imply the ability to express relevant ideas and concepts in an academic environment that are necessary for optimal performance. The second source of bias is tied to the psychometric properties of

the tools used for the assessment. The properties of the instruments must be kept invariant in relation to the language in which the assessment is carried out, and in relation to the linguistic groups. Psychometrics has developed quantitative and qualitative methodology for the study of the psychometric equivalency that has been focused on the adaption of tests (Muñiz, Elosua, & Hambleton, 2013) and the detection of differential item functioning (DIF; Dorans & Holland, 1992). The third element of the triad, the sociolinguistic bias is related with the impact on the assessment of variables such as the process of linguistic normalization and/or attitudes toward the languages (Elosua, 2016). From a broad perspective of the concept of score validity, the three aspects, the psycholinguistics, psychometrics, and sociolinguistics impact the assessment process, and so the score interpretation can be compromised if the competencies related with the domain of the academic language do not reach a minimal level, the language is not normalized and/or differential item functioning is detected.

The analysis of psycholinguistic and psychometric bias requires considering the effect of receiving instruction in a language different from the first language. This question is mainly addressed in the context of migrant students but it also affects other situations: students enrolled in a linguistic model whose instruction language does not coincide with their first language. In the United States, for example, the researches that compare the performance between immigrant English-learning students (ESL; English as a Second Language), and non-immigrants unanimously report worse performance for the first group (Abedi, 2009; Abedi & Lord, 2001; Solano-Flores & Trumbull, 2003). In the assessment of mathematics competence the results are associated with an irrelevant linguistic complexity (psychometric bias). In relation to the equivalence between tests administered in different languages it is common to report differential item functioning that becomes pronounced for the non-Indo-European languages (Grisay, de Jong, Gebhardt, Berezner, & Halleux-Monseur, 2007) and for multilingual countries (Monseur & Halleux, 2009).

In Spanish language contact environments, Pifarré, Sanuy, Huguet y Vendrell (2003) concluded that students instructed in Catalan whose first language is Spanish, on average, got math scores lower than their peers speaking Catalan at home. The authors also found a significant relation between the domain of Catalan and the performance in math in study with second-year secondary aged students (Sanuy, Huguet, Pifarré & Vendrell, 2002). Similar results have been found with relation to Basque; there are performance differences in math in fourth-year elementary aged students whose main language is Spanish and receive instruction in Basque in comparison to students whose main language is Basque and receive instruction in Basque (Elosua & Egaña, 2017). In order to explain those results the psycholinguistic hypotheses related with the Cognitive academic language proficiency (CALP) is appealed to.

In this research framework, the aim of this work is to study in-depth the potential sources of bias that can affect the results of any educational assessment in contexts of linguistic diversity. The study focuses on the Basque autonomous community (BAC) and mathematics competence. Since 1979, Basque is, along with Spanish, an official language in the BAC. In 1983, a so-called decree of bilingualism designed three bilingual education models known as models A, B, and D; each defined by an increasing proportion of Basque as the language of instruction. In the three models, Spanish language and literature, Basque language and literature, and modern languages are taught primarily in the corresponding language. The models were defined as follows:

- Model A: All subjects except Basque are taught in Spanish. Basque is treated as another subject with the number of weekly class hours stipulated by the Education and Culture Department.
- Model B: Both Spanish and Basque are used to teach other subjects.

- Model D: All subjects, except the Spanish language, will be taught in Basque, treating Basque the same as any other subject.

The selection of the math competency is not arbitrary; although it is important to recognize that language is the communication tool between the evaluator and the evaluated, we chose mathematics because theoretically it has a lesser linguistic load and therefore less involvement in the study of the linguistic variables.

Based on the revised literature; we formulate two research questions:

- In the models where Basque is the teaching language, either the partial (Model B) or full (Model D), one expects that no significant relation exists between the first language and the instruction language; it is to say, the math performance will be similar between students independent of their first language (psycholinguistic perspective).
- Given that minimal linguistic competencies required for the assessment tool have been acquired for all secondary education students (basic linguistic homogeneity), the exam will not contain significant differential item functioning (psychometric perspective).

Method

Sample

The sample is defined by 2,783 second-year students enrolled in secondary educational centers in the Basque Autonomous Community that receiving public funds, simultaneously offer the linguistic models B and D. 45 centers are included, 27 of which are public, and 18 partially funded by the state. In total, 1,341 of the students are female, and 1,442 are male. They are all natives of the BAC, and their first language is Basque (N = 624) or Spanish (N = 2,159). The student distribution according the linguistic model indicates that 854 are enrolled in model B and the rest, 1,929, study in model D (see table 1).

Instrument

In the framework of the educational diagnostic assessments, the Basque institute for research and evaluation in education constructed a 24-item questionnaire to assess mathematical competence (ISEI-IVEI, 2012a). Mathematical competence is defined as the ability to use and relate numbers, their basic operations, symbols and forms of expression and mathematical reasoning, as much to produce and interpret information as to expand knowledge of quantitative and spacial aspects, and to resolve problems related to daily life and the working world. The curriculum for the development of mathematical competence is the same for the students of both models B and D. The test was administrated in Basque for all of the participants.

Variables

Two types of variables were specified; individual-level variables or first-level variables and school variables or second-level variables.

Individual-level variables. Five predictor variables were defined in this level: index of economic, social and cultural status (ESCS), gender, suitability and first language.

- ESCS is a composite indicator constructed from the information provided by the parents of the student participants in the diagnostic assessment that was obtained from measures related to the educational level of the parents, the social status of their occupations, and the family resources. The index is measured in a standardized scale (M = 0, SD = 1).
- Gender: is a widely used variable in educational assessment. Studies about mathematical competence and/or mathematical performance and gender find a significant relationship between both variables with higher averages for men than for women (OECD, 2016).
- Suitability: reflects if the student is enrolled in the academic course that is theoretically adequate for their year of birth.

- First language. The first language is Spanish or Basque. The variable that is informed by the families, identifies the language regularly used in the household.

School variables. Three variables related to the characteristics of the educational center were delimited: Ownership, school socioeconomic index ($School_{ESCS}$), and linguistic model.

- $School_{ESCS}$. On the individual ESCS values, the $School_{ESCS}$ for each of the participating centers was estimated. It can be expected that as the resources provided for the students increase, the students' medium performance improves (Coleman et al., 1966; Willms, 2006).
- Linguistic model. The students in a school can receive instruction in the linguistic model B or in the linguistic model D.
- Ownership: indicates if the center is public or not. 1,725 students study in public centers and 1,058 in centers that are partially funded by the state.

Procedure and Analysis

The test administration was accomplished through external personnel at the schools during school hours. Being a diagnostic test, the data are censal.

The analysis integrates a psychometric perspective with a predictive approach based on the estimation of lineal mixed models (Goldstein, 2003). In the first approximation, the equivalence between the competence tests was analyzed according to the first language and the educational model. To achieve this, once the internal consistency coefficients were estimated, the percentages of variance associated with the first component extracted from an analysis of principal components were evaluated, and finally the presence of differential item functioning was studied. Two methods were applied to detect DIF: the Mantel-Haenszel statistic and standardization. For the classification of the severity of the differential item functioning, the criterion of the *Educational Testing Service* was used, which defines three levels: A (severe), B (moderate), and C (negligible).

For the study of the competence, in addition to the commonly used variables in the framework of the educational assessment as predictors of competence (ESCS; gender, ownership of the center, suitability) individual-level linguistic variables (first language) and second-level linguistic variables (educational model) were included. For all fixed effects, a dummy code was utilized. The levels of reference were respectively Spanish as the first language, model B as the linguistic model, the public network opposite the network partially funded by the state, female for the gender variable, and a student that is not enrolled in his theoretical school year. After the estimation of the mixed model, the least mean squares (lsm) were estimated. The *lmer* function of the *lme4* package was used for the estimation of the multilevel model (Bates, Maechler, Bolker & Walker 2015).

Results

Preliminary Analysis

In the response mechanization process, there was no considered difference between a missing value and an unreached item. Of the 1597 students, 57.02% responded to all of the test items. The percentage of missing values for each of the first 22 items ranged from 1% of the third item to 6% of item 19. The last two items of the test are associated with no responses of 8% and 20% respectively.

Of the distribution of students by educational models it can be concluded that families opt for model D by a greater margin. The percentage of students with Basque as the family language is 22.4% (N = 624); however, the number of enrolled students in model D, is superior to this number reaching 30.71% (N = 855). The relation between both variables is not independent (χ^2 (1) = 41.01; $p < .01$). Of all students whose first language is Basque, 79.80% (N = 498) choose model D as the instruction model; and this percentage is 66.2% (N = 1431) for those students who speak Spanish in their homes.

Except in the educational model variable, in which there are no observed differences in the math competency between models B and D

(Table 1), the differences related to the rest of the variables are significant ($p < .05$). The students from partially funded schools get a better math average (MA = 13.34) than those studying in public schools (MA = 12.72); the performance of the boys (MA = 13.15) is higher than that of the girls (MA = 12.74); and those students whose first language is Basque have a higher average (MA =14.13) than their native Spanish speaking counterparts (MA = 12. 62). The practical significance of these differences are evaluated with Cohen's *d*, that showed the higher size of the effect is associated with the first language variable.

Table 1 – *Sample, Descriptive Statistics and Differences in Math Competencies*

	N	Math Average	Contrast	*p*	*Cohen's d*
Educational Model					
Model B	854	12.83 (5.0)			
Model D	1929	13.01 (4.64)	$t(1530) = -0.88$	.37	-0.03
First Language					
Basque	624	14.13 (4.69)			
Spanish	2159	12.62 (4.72)	$t(1016) = 7.09$	< .01	-0.32
Gender					
Female	1341	12.74 (4.71)			
Male	1442	13.15 (4.79)	$t(2772) = -2.28$	.02	0.08
School type					
Public	1725	12.72 (4.75)			
Private	1058	13.34 (4.73)	$t(2244) = -3.33$	< .01	-0.12

Note: Between parentheses are standard deviations and degrees of freedom

Psychometric Equivalency according to First Language

The reliability coefficients estimated by the Cronbach's alpha on the tetrachoric correlation matrix were 0.87 and 0.88 for the student samples whose first language is Spanish and those speaking Basque regularly and prominently. For the first group, the variance percentage associated with the first component is 27%, and in the second group the estimated value is 26.5%. In both cases, the tests overcome the habitual cut-point for the determination of unidimensionality.

None of the DIF detection procedures concluded the presence of items that heavily favor either of the groups (Table 2). Only the item number 3 showed DIF favoring the native Basque speaking students. The indexes of difficulty for that item were .90 and .81 with much higher values for students that speak Basque in their natural environment. The discrimination indexes respectively were .23 and .39. For a better understanding of the behavior of the items the conditional item means related to the estimated competency levels were drawn (Figure 1). In the graphic is shown that item 3 barely discriminates on one of the groups, and that the biggest differences are found in the lowest levels of competency. For the rest of the items, the response averages or difficulty indexes for each competence level were similar independent of the first language. An item is shown as example of the patterns of the observed responses.

Table 2 – *Differential Item Functioning*

	First Language				Educational Model			
Item	**MH**	**Type**	**SD**	**Type**	**MH**	**Type**	**SD**	**Type**
1	0.97	A	0.00	A	1.15	A	-0.03	A
2	1.09	A	-0.01	A	1.13	A	-0.02	A
3	1.62	B	-0.06	B	0.86	A	0.01	A
4	0.95	A	0.02	A	0.94	A	0.02	A
5	0.90	A	0.00	A	0.93	A	0.01	A
6	1.01	A	0.00	A	0.98	A	0.01	A
7	0.77	A	0.06	A	1.05	A	-0.02	A
8	0.91	A	0.04	A	0.83	A	0.03	A
9	1.00	A	0.01	A	0.96	A	0.01	A
10	0.98	A	0.00	A	0.96	A	0.01	A
11	0.89	A	0.03	A	1.15	A	-0.04	A
12	1.04	A	-0.00	A	1.09	A	-0.03	A
13	1.00	A	-0.01	A	0.97	A	0.00	A
14	0.88	A	0.03	A	1.20	A	-0.03	A
15	0.97	A	0.01	A	0.77	A	0.04	A
16	0.96	A	0.01	A	1.01	A	-0.00	A
17	0.97	A	0.01	A	1.15	A	-0.03	A
18	1.14	A	-0.02	A	1.00	A	0.00	A
19	0.94	A	0.01	A	1.12	A	-0.02	A
20	0.95	A	0.01	A	1.02	A	-0.00	A
21	1.24	A	-0.03	A	0.87	A	0.02	A
22	1.09	A	-0.01	A	1.09	A	-0.02	A
23	0.99	A	0.02	A	1.11	A	-0.03	A
24	0.98	A	0.01	A	1.15	A	-0.05	A

Note: MH. Alpha Statistic of Mantel-Haenszel. SD=Standard Deviation.

Psychometric Equivalency based on the Educational Model

The reliability coefficients estimated by the Cronbach's alpha on the tetrachoric correlation matrix were 0.89 and 0.87 for the students of models B and D, respectively. In model B, the variance associated with the first component was 28.5%, and in model D the unidimensionality indicator was 26.6%.

The analysis of the differential item functioning in relation to the linguistic model rejected the presence of DIF; none of the questions reached the specific criterion.

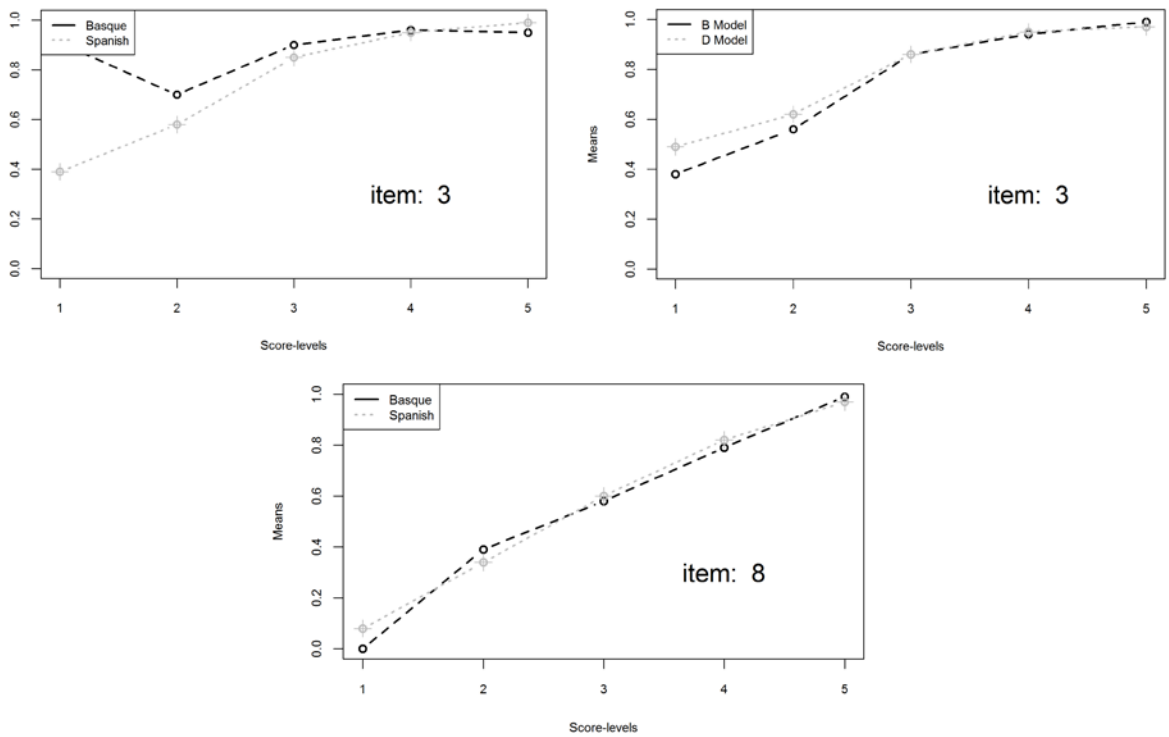

Figure 1 – Differential Item Functioning according home language and educational model. Conditional Difficulty Indexes.

Multilevel Predictive Model

In the first place the null model was estimated and it generated a residual variance of 20.51. The model that includes fixed effects variables such as gender, ESCS, School$_{ESCS}$, school type, linguistic model, first language and suitability was estimated by the maximum restricted likelihood (REML) method (Table 3). The intraclass correlation for this model was .065, which indicates that 6.5% of the variability between scores is associated with the characteristics of the schools.

Table 3 – *Summary of the Predictive Model: Coefficients (β), Standard Error (Se(β)), t-values and significance* (*p*).

Fixed Effects	**β**	**Se(β).**	***t***	***p***
Intercept	10.90	.43	25.29	<.01
ISEC	0.98	.09	10.46	<.01
Gender	0.54	.07	3.34	<.01
Educational Network	-0.42	.42	-0.97	.33
School$_{ESCS}$	1.45	.57	2.55	.01
Educational Model	-0.39	.23	-1.67	.10
First Language	-0.50	.22	-2.23	.02
Suitability	3.04	.24	12.28	<.01
Random Effects		Variance	SD	
School		1.21	1.10	
Residual		17.38	4.16	

As expected, the Socioeconomic and Cultural Index (ESCS) showed a significant effect on the individual level (β = 0.98; SE = 0.09; $t(2550) = 10.46$; $p < .01$), and on the school level (β = 1.45; Se = 0.57; $t(42) = 2.55$; $p = .01$). The parameter related with the gender also confirmed previous investigation results, that show a better performance by men in comparison to women (β = 0.54; Se = 0.16; $t(2550) = 3.34$; $p < .01$). The school system, semi-public or public, did not show differences (β = -0.42; Se = 0.42; $t(42) = -0.97$; $p = .33$). As for the first language, the results showed differences in favor of the group whose first language is Basque (β = -0.42; Se = 0.42; $t(42) = -0.97$; $p = .33$). The effects related with the linguistic model were not significant (β = -0.39; Se = 0.23; $t(44) = -1.67$; $p = .10$), showing that the estimated averages in math competencies are similar for students enrolled in models B and D in schools that offer both options. Finally, the students that study in the level theoretically corresponding to their age obtain better averages than their counterparts (β = 3.04; Se = 0.24; $t(2550) = 12.28$; $p = <.01$).

After the model estimation the least square means or adjusted means in math competencies were estimated; in other words, the mean values considering the fixed effects (Table 4). The means related with the educational model ($M_{modelB} = 12.32$; $M_{modelD} = 11.93$; $t(36) = 1.67$; $p = .10$) or the school system ($M_{public} = 12.34$; $M_{private} = 11.92$; $t(39) = 0.97$; $p = .33$), are not statistically significant. In relation to the first language, the averages are superior in the group of students whose first language coincides with the instructional language, Basque (M = 12.38) over those that combine two languages (M = 11.88); the difference is significant ($t(2358) = 2.23$; $p = .02$).

Results show that the highest competence levels are associated with male students from families with high socioeconomic levels whose first language is Basque and reach the level corresponding to their age; independently of the educational system or linguistic model.

Table 4 – *Adjusted means in Math Competency*

	Exact Average	Contrast (*t*)	*p*
Educational Model			
Model B	12.32 (.26)		
Model D	11.93 (.24)	1.67	.10
First Language			
Basque	12.38 (.28)		
Spanish	11.88 (.22)	2.23	.02
Gender			
Female	11.85 (.24)		
Male	12.40 (.23)	-3.34	< .01
System			
Public	12.34 (.29)		
Semi-public	11.92 (.33)	0.97	0.33

Note: In parentheses the standard errors of the exact averages.

Conclusions

Educational assessment is a key tool in improving the quality of education (Tiana-Ferrer, 2018). As a tool of improvement, educational assessment in contexts of linguistic diversity must respond to specific issues derived from the plurality of the population. Recognizing linguistic diversity as common and enriching, equity in the assessment process demands the consideration of aspects related to the psychometric characteristics of the tests, the development of linguistic competence, or the familiar linguistic context of the student. The comparability between scores and therefore the quality of the assessment process demands

the examination of different sources of score variation.

This work intends to delve into the problem of comparability in these circumstances. With reference to the first language and its impact on the assessment results, the psychometric equivalence of the tests based on the first language and the educational model has been concluded. Only one item has shown moderate differential item functioning that affects the lowest levels of competence.

In reference to the predictive model, as students are assembled in greater units, educational centers, a methodological approach was utilized that enables the estimation of fixed and random effects. The results show that the variance or variability in mathematical competence explained by the characteristics of the educational centers is only 6.5%. A high value in this index, the intraclass correlation coefficient, would be interpreted in terms of close relation between the centers and mathematical competence, which wouldn't be desirable from the educational equity point of view. However, it is necessary to explain this datum. The inter-center variability estimated in the diagnostic assessment of mathematical competence with the complete population of students registered in the Basque Country was 16.9% (ISEI-IVEI, 2012a). In this study the estimated value refers to a subset of 330 schools that form part of the Basque educational network: only those that offer two linguistic models concurrently, models B and D.

Emphasizing the effect of the characteristics of the schools, it is noted that if the arithmetic means and the adjusted means are compared, the difference by ownership is diluted. This is to say that, looking at the analyzed data, once the compositional effect is taken away, there are no differences in the centers added values. This circumstance can be interpreted as admitting that the difference between the arithmetic means does not come from the quality of the education offered from the public network and from the network partially funded by the state, since the averages of both are equalized when discounting the effects of the tuition. The origin of the differences could be situated in the fact that at the time of choosing the center/network, the families with higher ESCS indexes prefer to choose semi-public schools.

The first-level variables related to the individual or family characteristics confirm: a) the relation between ESCS and mathematical competence, b) the impact of unsuitability of the student on performance, c) the differences of mathematical competence based on gender, and d) the negative effect of the differences between the first language and the instruction language. ESCS, gender, and suitability have been widely debated inside the educational assessment and the results on its effects are in agreement (OECD, 2016). Regarding the impact of the first language on academic performance in contexts of linguistic diversity, the results contravene the initial hypothesis and reinforce the studies that show the relation between both variables.

The results strengthen the need to support the acquisition of linguistic competencies of students whose first language is different from the instruction language, a conclusion that was also deduced from the research carried out in Cataluña (Sanuy, Huguet, Pifarré & Vendrell, 2002). The basic hypothesis that guarantees equity in the assessment process –except the evaluation of oral and related linguistic competencies –is that differences between those evaluated do not exist in the linguistic level required for the optimum test performance. This basic principle of minimum linguistic homogeneity can be violated in multilingual contexts and as a result is converted into a source of irrelevant variance.

Finding differences in mathematical performance between students of secondary schooling that are native students studying in models B and D seems to question the acquisition process of CALP. However, just as Cummins (2000) points out, academic knowledge of a language is not equivalent to the knowledge of using the language, rather it makes reference to understanding and utilizing the specific type of language used in educational contexts. This reflection is the

origin of a debate that derives toward the effect of the test administration´s language (Abedi & Lord, 2001; Solano-Flores, 2016), and that seems to conclude that in multilingual contexts the language that maximizes performance is the first language (Elosua & Mujika, 2014; ISEI-IVEI, 2012b).

Extending the matter of equity to those schools that share more than one linguistic model, the results have shown that differences do not exist with respect to this variable. Although the objective of the study was not the evaluation of linguistic models, in the research development it was important to consider its possible effect on the results. In relation to the characteristics of the population studied –students in required secondary schooling registered in centers that conjointly offer models B and D –the educational model factor was not significant. This conclusion is highly satisfactory from the perspective of educational equity. However, it is important to point out that it would be interesting to extend the study and include centers that offer model A, characterized by the use of Spanish in teaching as an instruction language.

In conclusion, the comparability between scores in linguistic diversity contexts demands the conjunction of different areas of knowledge. The comparability becomes even more difficult when an educational system has distinct models or programs available, that can also be offered by the same center. It is an issue with countless political, educational, and social connotations, addressed in an elusive and confusing manner by the specialized bibliography. Placing the educational databases for public disposal could help to delve into a topic that is just as controversial as it is fundamental.

References

Abedi, J. (2009). Validity of Assessment for English Language Learning Students in a National/International Context. *Estudios sobre Educación, 16*, 167-183. doi: https://doi.org/10.1207/S15324818AME1403_2

Abedi, J., & Lord, C. (2001). The language factor in mathematics. *Applied Measurement in Education, 14*, 219-234.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *671*, 1-48. doi: https://doi.org/10.18637/jss.v067.i01

Coleman, J., Campbell, E. Q., Hobson, C. J., McPartland, J., Mod A. M., Weinfeld, F. D., & York, R. L.(1996). *Equality of Educational Opportunity*. Washington, DC: US Department of Health, Education and Welfare.

Cummins, J. (1981). Four misconceptions about language proficiency in bilingual education. *NABE Journal, 5*, 31-45. doi: https://doi.org/10.1080/08855072.1981.10668409

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters, Ltd. doi: https://doi.org/10.1080/15235882.2001.10162800

Dorans, N.J., & Holland, P.W. (1992). *DIF detection and description: Mantel- Haenszel and standardization*. (RR-92-10). Princeton, NJ: Educational Testing Service.

Elosua, P. (2016). Minority Language Revitalization and Educational Assessment. Do Language-related factors impact perfomance?. *Journal of Sociolinguistics, 20*, 212-228. doi: https://doi.org/10.1111/josl.12176

Elosua, P., & Egaña, M. (2017). Socioeconomic patterns and educational performance based on language-of-instruction models in the Basque Autonomous Community. *Journal of Multilingual and Multicultural Development, 38*, 686–698. doi: https://doi.org/10.1080/01434632.2016.1224239

Elosua, P., & Hambleton, R. K. (2018). Psychological and Educational Test Score Comparability Across Language and Cultural Groups in the Presence of Item Bias. *Revista*

de Psicología y Educación 13, 23-32. doi: https://doi.org/10.23923/rpye2018.01.155

Elosua, P., & Mujika, J. (2014). Impact of family language and testing language on reading performance in a bilingual educational context. *Psicothema, 26*, 328-335. doi: https://doi.org/10.7334/psicothema2013.344

Grisay, A., de Jong, J.H.A.L., Gebhardt, E-. Berezner, A., & Halleux-Monseur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement, 8*, 249-266.

Goldstein, H. 2003. *Multilevel Statistical Models*. Third Edition. London: Edward Arnold.

ISEI-IVEI. (2012a). *Evaluación Diagnóstica. Informe de Resultados*. Bilbao: ISEI-IVEI.

ISEI-IVEI. (2012b). *PISA-L*. Bilbao: ISEI-IVEI.

Monseur, C., & Halleux, B. (2009). *Translation and verification outcomes: National versions quality*. In OECD technical report (pp. 96-104). Paris: Organisation for Economic Cooperation and Development.

Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de tests: Segunda edición [Guidelines for the translation and adaptation of tests: Second edition]. *Psicothema, 25*, 149-155. doi: https://doi.org/0.7334/psicothema2013.24

OECD. (2016). *PISA 2015 Results: Excellence and Equity in Education.* Paris: OECD Publishing, Paris. doi: https://doi.org/10.1787/9789264267510-en

Pifarré, M., Sanuy, J., Huguet, A., & Vendrell, C. (2003). Rendimiento matemático en contextos bilingües. Análisis de la incidencia de algunas variables del contexto socio-educativo [Mathematical performance in bilingual contexts. Analysis of the incidence of some variables of the socio-educational context]. *Revista de Investigación Educativa*, 21,183-199.

Sanuy, J., Huguet, A., Pifarré, M., & Vendrell, C. (2002). Relaciones entre el conocimento del catalán, el castellano y el rendimiento matemático. Un estudio empírico [Relations between the knowledge of Catalan, Spanish and mathematical performance. An empirical study]. *Revista de Educación*, *328*, 383-394.

Solano-Flores, G. (2016). *Assessing English Language Learners*. New York: Routledge. doi: https://doi.org/10.4324/9780203521953

Solano-Flores, G., & Trumbul, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English Language Learners. *Educational Researcher, 32*, 3-13. doi: https://doi.org/10.3102/0013189X032002003

Tiana-Ferrer, A. (2018). Treinta años de evaluación de centros educativos `en España [Thirty years of evaluation of educational centers in Spain]. *Educación XXI*, *21*(2), 17-36.

UNESCO (2003). *Education in a multilingual world. UNESCO Education Position Paper*. Paris: UNESCO. Retrieved from http://unesdoc.unesco.org/images/0012/001297/129728e.pdf

Willms, J. (2006). *Learning Divides: Ten policy Questions About the Performance and Equity of Schools and Schooling Systems*. Montreal: UNESCO Institute for Statistics.

Zalbide, M., & Cenoz, J. (2008). Bilingual Education in the Basque Autonomous Community: Achievements and Challenges. *Language, Culture and Curriculum*, *21*, 5-20. doi: https://doi.org/10.2167/lcc339.0

Acknowledgments

Research partially funded by the UPV / EHU GIU 15/24, GIU18 / 88

Assistant translators:

Alyson Ellison and Nathan R. Young

Authors / Autores	To know more / Saber más
Elosua, Paula (paula.elosua@ehu.es) Doctor in Psychology in the Department of Social Psychology and Methodology in Behavioral Sciences at the University of the Basque Country. Psychology Faculty. Mail address: Facultad de Psicología. Avda. Tolosa, nº 70. 20018-San Sebastián (Guipúzcoa, Spain).	ORCID 0000-0002-9651-9100 Google académico ResearchGate

Revista **EL**ectrónica de **I**nvestigación y **EV**aluación **E**ducativa

E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]