

RESEARCH ARTICLE

Guatemalan Pupils' Exposure to Spanish and Educational Achievement: A Study across Ethnic Groups and Socioeconomic Status in Urban and Rural Schools

ALVARO M. FORTIN MORALES

Universidad del Valle de Guatemala, Guatemala and Tilburg University, the Netherlands

FONS J. R. VAN DE VIJVER

*Tilburg University, the Netherlands, North-West University, South Africa
and University of Queensland, Australia*

YPE H. POORTINGA

Tilburg University, the Netherlands

Abstract

We investigated the relationship of exposure to Spanish and socioeconomic status (SES) with reading and math standardized test scores among third and sixth grade pupils in Guatemala. The results of a multi-group structural equation model and multivariate analysis of covariance led us to conclude that SES and exposure to Spanish are associated to achievement across grades, ethnicities, and area of residence. Exposure to Spanish is more relevant in lower grades whereas area of residence is a consistent predictor across grades. Analysis of equivalence for the various groups suggests a set of consistent relationships across ethno-linguistic groups. Our results confirm previous research indicating that SES is a positive predictor of school achievement and that the compatibility of the pupil's familiarity with the testing language is positively related to achievement results. We recommend that linguistic competency heterogeneity in terms of competency level in each language be considered and that bilingual interventions be developed, as opposed to adopting models that assume homogeneous linguistic ability across ethnicities.

Keywords

Maya, indigenous, Ladino, achievement, education, Guatemala, language mastery

Guatemala is a multi-ethnic and multi-lingual society. Spanish is the official and dominant language in the education system. Efforts have been made to provide a more equitable treatment of all languages and sociodemographic shifts have led to increased interactions between linguistic groups where Spanish is the common language. This has created a national population with huge individual differences in proficiency in Spanish. In this paper we examine effects of differential exposure to Spanish and

socioeconomic status (SES) on pupils' school performance. We used scores of standardized tests for math and reading across ethnic groups, while taking into consideration the area of residence (rural or urban). The results should be of immediate interest for the Guatemalan educational system and also offer interesting inputs for a wider context since most countries where pupils are schooled in multi-lingual environments, will also have to negotiate between languages. Understanding these relationships should inform educational

policy in Guatemala and presumably other multi-lingual countries and serve as a stepping stone for further research into differential levels of language proficiency among pupils.

Ethnicity, Language, and Socioeconomic Status

Ethno-linguistic factors are difficult to separate from other relevant contextual factors in Guatemala. Previous research has consistently reported relationships between area of residence (urban/rural), SES and school achievement (Álvarez & Schiefelbein, 2007; Beckett & Pebley, 2003; Esquivel Villegas, 2006; Moreno Grajeda, 2009, 2012; Pellecer Rivera, Saz Choxin, & Santos Solares, 2011; Treviño Villareal, 2006). These are factors that in Guatemala are confounded with each other and with ethnic background. Of the total population, 40% to 50% is estimated to belong to one of the Mayan ethnic groups who mostly live in rural areas (Población en Guatemala (demografía), n.d.; World Factbook, 2007). Mayas are also, on average, less affluent and have lower school-attendance records than the mainstream group (i.e., Mestizos, also locally called Ladinos), and the income of families where the head of the household is not fluent in Spanish is usually lower (Esquivel Villegas, 2006; INE, 2010).

Findings from national school assessments in Guatemala show that SES of the individual pupil is the strongest predictor of achievement. It is not, however, a straightforward indicator; it loses predictive value as the average SES of pupils in the school increases (Moreno Grajeda, 2012). In schools with a more affluent population, performance of pupils is more homogeneous irrespective of SES variability within classrooms, while in poorer schools, pupils of lower SES obtain lower scores (Moreno Grajeda, 2012). Mayan pupils tend to belong to the lower SES segment of the poorer schools in the country.

Ethnicity and Enrolment Rates

According to the national education census, gross enrolment rates for third grade between 2004 and 2007 oscillated between 112% and

114% and net enrolment rates oscillated between 34% and 39% (Sistema Nacional de Indicadores Educativos, n.d.). By 2014 the gross enrolment rate was 99.1% and net enrolment rate 39.2% (Sistema Nacional de Indicadores Educativos, n.d.). Gross enrolment rate is estimated by dividing the number of enrolled students irrespective of age over the number of students in the general population of an age appropriate for the grade level. Net enrolment rate is estimated by dividing the number of students enrolled that are age-appropriate for the grade over the age-appropriate general population. In 2006, the year from which the data for this study is drawn, close to 75% of pupils in third grade were over-aged and around 60% of children who should have been enrolled were either in a grade not appropriate for their age or out of school. The data for 2014 indicates that perhaps the proportion of over-aged pupils has decreased, but the rate of age-appropriate children has shifted only slightly. Sixth grade data indicates that for the same period discussed above, 2004 to 2007, gross enrolment rates increased from 71% to 78% and net enrolment rates increased from 22% to 27%, while in 2014 the gross enrolment rate was 86.6% and the net rate was 32.8% (Sistema Nacional de Indicadores Educativos, n.d.). In other words, by the sixth grade the percentage of over-age pupils within the period of 2004–2007 dropped to around 50%, but the number of children that should have been in sixth grade and were in a grade not appropriate for their age or out of school climbed to around 75%. Considering the 2014 indicators in sixth grade, it seems that the over-age students increased with time, but there was also an increase in enrolment of age-appropriate children.

The promotion rate of students who completed second grade in 2005 and could move to third grade in 2006 was 84% and those completing third grade in 2006 was 82%. In contrast, the promotion rate of students completing fifth grade in 2005 who could move to sixth grade in 2006 and the rate for those completing sixth grade in 2006 was 91%. This suggests that the characteristics of children in third and sixth grade were somewhat different. The third grade group contained a larger proportion of the population

of children of the corresponding age (ideally, 8.5 to 9.5 years of age) and a larger intake of children overall, thereby constituting a more accurate representation of the country's population.

Of the 2,269,520 children enrolled in 2006, the year of the data collection for this study, only 38% were indigenous (including Maya, Garífuna, and Xinka) (Anuario Estadístico de Educación, 2006). Of the total school population, 70% attended rural schools, but this proportion climbed to 85% when considering only indigenous pupils. Promotion rates in that year were 82% for the rural area and 89% for the urban area. The strong link between area of residence and ethnicity is likely to have contributed to create heterogeneity in school populations across grades.

Ethnicity, Language, and Education Delivery

Historically, the linguistic diversity in Guatemala has created significant challenges for the delivery of educational services, and population mobility further contributes to these. Demographic shifts caused by diverse socioeconomic and historical factors have gradually reduced the isolation of ethnic groups, changing the linguistic landscape of the country in ways that have not been recorded accurately. Changes in educational policy have not always been capable of responding to those shifts simultaneously to ensure quality of education.

The Ministry of Education of Guatemala has designated a number of schools to deliver bilingual education to pupils in first to third grade in areas historically linked to a particular ethno-linguistic group. The reports on the success of this program usually have been positive in terms of enrolment and dropout rates, but inconsistent in terms of adherence to the use of the language in the classroom and the achievement of pupils (Enge & Chesterfield, 1996; Patrinos & Velez, 2009; Rubio, 2004). It is unclear whether this is due to a mismatch of teacher and pupil language, or that the teacher needs to translate the lesson or curriculum materials because pupils are not fluent in the language assumed to dominate the communities served by the school, or a

combination of these and perhaps other factors.

Guatemala's national assessment system tests pupils' achievement in math and reading as part of an overall strategy to monitor quality of education. Although there is general agreement that linguistic factors can affect outcomes, these are seldom tested. The attempts made in Guatemala focus on the "monolingual – bilingual" dichotomy (bilingual in Spanish and a Mayan language), as if there were no difference in levels of competency, assuming that the pupil is fluent in Spanish when monolingual and most fluent in the Mayan language when bilingual. For example, pupils belonging to a Mayan ethnic group and attending schools with bilingual teachers have on occasion been administered a test in a Mayan language designed to test math or reading among proficient speakers of that language. However, there is no additional information of the oral fluency the pupils possess in that language. As a result, even when pupils are Maya, they might have been tested in a language they are not fluent in. This is an important issue because the difference between the pupils' language of choice and the standard language in which they are tested can have a negative impact on test performance (Hambleton, Merenda, & Spielberger, 2005; Treviño Villareal, 2006).

Guatemala is home to 22 languages. The mother language (L1) for many children is not Spanish and the degree to which they have contact with Spanish varies. Some pupils who identify with the Mayan group do not speak a Mayan language. On the other hand, some pupils who identify with the Ladino group speak a Mayan language or have parents who speak a Mayan language. The extent to which Spanish dominates over other local languages or the extent to which Spanish and Mayan languages hybridize each other also seems to differ across geographical locations. As a result, the mastery of Spanish is not homogenous across the population as a whole or within ethnic groups, making it difficult to ensure a good adjustment between teacher's and pupil's mastery of the language of schooling.

Not all bilingual communities agree with bilingual education and the attitude of the community may have an impact on the

consistency of the language in which the lessons are delivered and the methods used to deliver the instruction (Flores Reyes & Villegas de Sanders, 2009). Shifts in the background of the pupils as the grade level progresses change the composition of the classroom in more or less systematic ways, where pupils who reach higher grades tend to come from wealthier homes in which Spanish is spoken more frequently. The ethnic composition of the classroom may also have an impact. Evidence from classroom observations indicates that teachers and pupils initiate more interactions with those of their own ethnicity, thus creating different opportunities to learn as a function of the teacher-pupil ethnic composition (Chesterfield, Enge, & Rubio, 2002; Correa-Chávez & Rogoff, 2009), although it is not completely clear whether this is due to cultural or linguistic factors. This highlights the relevance of studying both ethnic and linguistic issues and their relationship with schooling and achievement.

Present Study

This study investigated the relationship of exposure to Spanish with pupils' school performance, controlling for socioeconomic status, and taking into consideration area of residence (rural or urban) and grade level. We wanted to find out whether pupils who have greater exposure to Spanish differed from pupils who come from bilingual homes and have less opportunity to communicate in the language of the standardized tests. As mentioned, Guatemala is linguistically heterogeneous and language, socioeconomic status and ethnicity interact, and are associated with exclusion. We also decided to explore whether gender is associated with differences in achievement results, because being female is also associated with exclusion from opportunities to achieve further development, such as education. Guatemala's Gender Development Index in 2013 was 0.52, placing the country in rank 112 (out of 209 countries) (UNDP, 2014). Having a clearer understanding of these interactions should help make better decisions regarding educational policy.

To answer our main question, whether there are differences in achievement associated with differential exposure to Spanish, we explored the national standardized educational assessments seeking sets of data that contain information on language proficiency. We were unable to find measures of linguistic mastery in either Spanish or the Mayan languages for students in the elementary grades. Testing for the level of mastery for all languages spoken in Guatemala would not be possible at this time, but measuring linguistic performance in the language of the tests would help understand observed differences in performance. At present there are no measures of fluency in Spanish of pupils and developing them would require a significant investment from the Ministry of Education.

The official policy for bilingual education is to deliver lessons in the Mayan language for indigenous children during the basic grades of elementary education (first three years for children, ideally 6.5 to 9.5 years of age). This, it is believed, will ensure better performance when pupils graduate from elementary education (in sixth grade, with children ideally ranging from 11.5 to 12.5 years). The national assessment system tests pupils concluding the basic cycle (third grade) and concluding the full elementary level (sixth grade). This happens every few years; the time between elementary level testing varies depending on approval from the incumbent Minister of Education.

In the record of the 2006 national assessment, we found pupil reports on the level of exposure they have had to Spanish. We decided to use this information to check whether the degree of exposure to Spanish among pupils who identify (or not) as Mayan and who belong to different socioeconomic levels relates to their educational achievement as measured by their performance in math and reading standardized tests. Results indicative of a pattern of relationships with socioeconomic status, ethnicity and area of residence (rural or urban) would justify an investment in developing the tests and mechanisms to assess proficiency in Spanish as a second language. We analyzed data from grades

three and six, to check whether trends replicate across levels and chose the set from 2006 because it contains equivalent information on exposure to Spanish and socioeconomic status across grade levels. We also used the same data set to check for differences in achievement associated with exposure to Spanish, SES, area, and gender.

Method

Participants

Participants were Guatemalan third and sixth grade pupils from government schools sampled to conduct the national reading and math 2006 assessment (about 89% of students enrolled in the elementary level attend a government school). Probability samples were drawn from each of the 22 Departments into which the country is divided (Ministry of Education/DIGEDUCA, n.d.). The sample for the Department of Guatemala, where the capital city is located, was split in two: One sample for the metropolitan area and one for the remainder of the Department. We chose data from the 2006 assessment because it was the only one of the sets available for research that contained information on exposure to Spanish, language preference, SES, ethnicity, and gender for all groups and grade levels.

The total sample for third grade was composed of 20,925 pupils, 50% males and 50% females, with a mean age of 10.16 years ($SD = 1.34$). The sample was divided into students who identified as Mayan (the majority indigenous population of Guatemala composed of some 20 distinct ethnicities) and Ladino (or mestizo). Of the pupils, 34% identified as Mayan and the others as Ladino; of the pupils who identified as Mayan, 22%

were highly over-aged for the grade level (12 years of age or older) as opposed to 13% of the non-Mayan group. The sample was also divided according to the municipal classification of the community served by the school as rural or urban. Of the pupils in the sample, 41% came from urban settings and 59% from rural areas; 68% of pupils who identified as Mayan were living in rural areas as opposed to 55% of Ladino pupils.

The total sample size for sixth grade was 15,227, 50% male and 50% females, mean age of 13.10 years ($SD = 1.28$), with 31.6% of pupils identified as Mayan and the rest as Ladino. Twenty percent of the pupils who identified as Mayan were severely over-aged for the grade level (15 years of age or older) as opposed to 11% of the non-Mayan group; 50% of pupils in the sample come from urban settings and 50% from rural areas; 57% of pupils who identified as Mayan live in rural areas as opposed to 47% of Ladino pupils.

There were a number of students who completed the reading test but did not complete the math tests or were administered, for either reading or math, a test form that could not be anchored to the remaining forms (see below). In third grade 13,625 pupils had a reading score and 15,406 pupils had a math score. In sixth grade 10,150 pupils had a reading score and 12,442 pupils had a math score. Little's MCAR test to check for missing reading and math data resulted in non-significant χ^2 value, 1.88 ($df = 2$, $p = .39$) for third grade and 1.27 ($df = 2$, $p = .53$) for sixth grade. Therefore, we used EM imputation estimations for reading and math scores, yielding a sample of 20,672 cases for third grade and 15,227 cases for sixth grade. Pupil numbers distributed by ethnicity, area of residence, and gender are presented in Table 1.

Table 1. *Numbers of Students by Gender, Area and Ethnicity*

Ethnicity	Area	Gender	3 rd Grade	6 th Grade
Maya	Urban	Male	1,053	899
		Female	1,206	1,189
	Rural	Male	2,481	1,469
		Female	2,300	1,249
Ladino	Urban	Male	2,958	2,703
		Female	3,216	2,815
	Rural	Male	3,825	2,520
		Female	3,633	2,383

Instruments, Questionnaire based measures

Languages exposure. There was only one version of the pupil questionnaire per grade with most questions common to third and sixth grade. Pupils reported their own and their parents' first language in three different questions (one referring to the pupil, one to the mother and one to the father), choosing from a list of the languages in use in Guatemala. They also reported what language they used to communicate with their father, mother, and siblings and indicated which language they used most frequently. The questions came with a list of languages from which the pupils could choose.

In total, seven items that are common across the two grades were scored dichotomously (coded as Spanish versus all other languages) and analyzed using Winsteps 3.66 (Linacre, 2008) for each grade separately. Cronbach's alpha was 0.91 for third grade and 0.93 for sixth grade. Rasch model fit statistics were used to identify response patterns not in line with (or predictable by) this model (Bond & Fox, 2007). We used the infit and outfit measures of the items as fit statistics (Bond & Fox, 2007). Rasch infit and outfit measures were considered to be under-fit when the mean square statistic was > 2.0 and over-fit when the mean square statistic was < 0.5 (as suggested by Bond & Fox, 2007). Since no item exhibited infit or outfit above 2.0 or below 0.5, we proceeded to estimate person Rasch measures for the scale, using all items.

Socioeconomic status (SES). Four items in the background questionnaire enquired whether the father and mother of the child knew how to read and write. A set of three multiple response questions asked about the materials of the floor and walls of the pupil's home where the options could be scored on an ascending scale (e.g., 1 point for walls made of reused tin or plastic, 2 points for prefabricated materials, 3 for wood, 4 for clay and brick, etc.). Further multiple choice questions asked about the water-source and the source of energy for illumination at home; a list of objects indicative of wealth from

which the pupil chose those owned by his or her family (e.g., computer, DVD player, and telephone). A yes/no question asked whether or not the pupil was living in a house with a sewer system.

Using these items, raw scores were estimated for seven variables: i) objects indicative of wealth at home; ii) floor materials; iii) wall materials; iv) roof materials; v) source of illumination; vi) water source; vii) literacy of parents. Cronbach's alpha for third grade was 0.56 and 0.57 for sixth grade. These seven scores were subjected to principal component analysis. Sampling of cases for the analysis was adequate (global KMO for third grade = .61; global KMO for sixth grade = .68). Correlations were large enough for the analysis (Bartlett's Test of Sphericity for third grade: $\chi^2(15) = 10245.22$, $p < .001$; Bartlett's Test of Sphericity for sixth grade $\chi^2(15) = 9923.27$, $p < .001$). A factor score was estimated for each pupil based on the one-factor solution.

Instruments, Educational achievement tests

There were several test forms for the reading and math tests, depending on subject and area (as described later in this section). All but one of the forms for each school subject shared common items with the remaining forms. There were no common items across grades. The tested contents were set according to specification tables based on Marzano's taxonomic framework. This taxonomy classifies thinking skills in one domain and three systems: knowledge domain, cognitive system, metacognitive system, and self-system (Marzano, 2006). The taxonomy was developed for pedagogical planning and can be used to develop tests. According to the model, the self-system decides when actions are to be taken, the metacognitive system sets the goals and monitors them, the cognitive system processes the necessary information and the knowledge domain provides the necessary content (Marzano, 2006). Guatemala's national assessment includes items on the knowledge domain and the cognitive system with four sub-systems

(knowledge retrieval, comprehension, analysis, and knowledge utilization).

Third grade reading assessment contents included assessing vocabulary and finding information in a text at the taxonomic domain of knowledge; extracting information from context, meaning of expressions or syllogisms, identifying the main character at the level of the system of comprehension; understanding the intent of the author, identifying the main topic and predicting events based on the information in the text at the level of analysis in the cognitive system (MINEDUC/DIGEDUCA, n.d.). Third grade math assessment contained assessment of natural numbers, measurements (length, volume, time), and Mayan numeric system at the taxonomic domain of knowledge; arithmetic at the level of comprehension in the cognitive domain; word problems, geometric principles, and measurements at the level of knowledge utilization in the cognitive system (MINEDUC/DIGEDUCA, n.d.).

Sixth grade reading assessment contained identification of details from context and chronological sequencing at the taxonomic level of knowledge; vocabulary, extracting vocabulary meaning from context, predicting events based on the content of the text, identifying main topic and inferring precise meaning at the level of utilization; identifying intent of the author, identifying main ideas and extracting conclusions at the level of comprehension; making generalizations and detecting similarities and differences at the level of analysis (MINEDUC/DIGEDUCA, n.d.). Sixth grade math assessment included contents on geometry and aggregation at the knowledge level; probability and analysis, percentages, measurements and word problems at the utilization level; arithmetic at the comprehension level (MINEDUC/DIGEDUCA, n.d.).

There were three forms of the third grade Spanish reading test, each consisting of 40 multiple-choice items with four response alternatives. Two of the forms shared 14 common items. Cronbach's alpha values for these two versions were 0.80 and 0.82. There were three forms of the sixth grade Spanish reading test, each consisting of 40 multiple-choice items with four response alternatives. Two of the forms shared 16 common items.

For these two forms Cronbach's alpha values were 0.82 and 0.81.

There were four forms of the third grade math test, each consisting of 30 multiple-choice items with four response alternatives. Three forms of the test shared 22 common items. Cronbach's alpha values were 0.82, 0.84 and 0.82. There were five forms of the sixth grade math test, each consisting of 40 multiple-choice items with four response alternatives. Four of the forms shared 16 common items. Cronbach's alpha values were 0.73, 0.77, 0.78 and 0.73.

Procedure

The technical report for 2006 provides a detailed description of the administration procedures (MINEDUC/DIGEDUCA, n.d.). The 2006 assessment took place in August to ensure that the majority of pupils were still attending school. The school year in Guatemala starts in January and finishes in October, but it is common for pupils to leave school prematurely by September.

The sample for 2006 was composed of 718 schools. The standardized assessment was administered by a team directly under the Directorate for Educational Assessment (DIGEDUCA) and not by the school teachers. This was done to prevent the loss of tests, decrease the possibility of corruption during administration, ensure greater homogeneity in administration and secure personnel that had been specifically trained for test administration. The teams were required to visit each of the 718 schools in a period of three to four weeks.

Each team consisted of two to four individuals. The team met with the teachers and the principal, and administered the tests to the children. In each grade the math test was completed first, then the reading test and finally the pupil questionnaire.

The different versions of a test were distributed systematically to minimize cheating. The instructions of the math and reading tests were read by the test administrator while the pupils followed in their own booklets. Two examples were made jointly and then the pupils worked on their own. No questions related to the content of the test were answered during the administration.

The pupils answered directly on the booklet (no answer sheets were used).

The administration of the questionnaire was guided. The administrator would read each question while the pupils read their own booklets; time was provided for each pupil to answer before moving to the following item. Any question that would help in clarifying the meaning of the items was answered to the whole group. The pupils answered directly on their booklets.

Statistical Analyses

Our analysis examined whether exposure to Spanish has a predictive value for educational achievement, acknowledging that socioeconomic status also influences these results. We were interested in checking for patterns of this relationship across ethnic groups due to their association with languages other than Spanish, and across grade levels given the shifts in the composition of the pupil population as schooling progresses. We used national math and reading tests to assess achievement.

Different reading and math test versions were anchored to a common scale using the Rasch model after item analysis had been conducted. Before anchoring, items of each version of the test and of each subject area (math and reading) were analyzed separately for (a) Differential Item Functioning (DIF) using logistic regression; (b) improvement of Cronbach's alpha after deletion of each item; and (c) Rasch's infit and outfit measures. Logistic regressions and Cronbach's alpha values were computed with SPSS 15.0. Rasch infit and outfit measures were estimated with Winsteps 3.66 (Linacre, 2008).

Bias appears when there are nuisance factors that cause the meaning of the assessment tools to vary across groups, preventing the direct comparison of scores (Poortinga, 1989). Differential Item Functioning or item bias occurs when individuals of the same standing in the construct have different probability of answering an item correctly (Van de Vijver, 2011; Van de Vijver & Tanzer, 2004). Biased educational testing can be unfavorable to lower-performing groups, particularly when

testing has some type of repercussion on the pupils' opportunities for further education (Camilli, 2006).

DIF was estimated using a logistic regression where item correct/incorrect response was the dependent variable and total test score was the predictor. Area of location of school (urban and rural areas) and the ethnicity of pupils were added as covariates in two separate steps to assess uniform and non-uniform DIF. An item was considered to show DIF when the significance level of the regression was 0.05 or less and the difference of the Nagelkerke R² between the first step of the analysis and the step where the covariates had been added was 0.035 or larger (as recommended by Sireci, 2011). No significant DIF was identified in any of the math or reading test forms. This was somewhat surprising because in a previous study (see Fortin Morales, van de Vijver, & Poortinga, 2013) for which we used similar populations, but other tests, we found substantial evidence of DIF. However, in that study, deletion of DIF items did not cause significant changes in the overall test results. Perhaps test development in Guatemala has improved in such a manner that item sets are better calibrated; it is also possible that school populations have become more homogenous as a result of increased enrolment. In any case, the current version of the test satisfied the demands of this study.

Deletion of items that are not consistent with the construct being tested should increase Cronbach's alpha. In this data set, the deletion of items produced negligible changes only (less than 0.01), which can be seen as another indicator of careful test development. We used the infit and outfit measures of the items as fit statistics, as suggested by Bond and Fox (2007). One item in the third grade math tests exhibited underfit for infit and outfit measures and one item for third grade math exhibited underfit only for outfit measures. These two items were common to all versions of the math test and were deleted before anchoring. No items were removed from the third grade reading test or the sixth grade tests.

Once we were satisfied with the achievement measures, we analyzed the relationship between pupils' achievement,

measured by math and reading scores, with their SES and their exposure to Spanish. We also checked whether this relationship is moderated by ethnicity and area of residence. To accomplish this, we fitted a multi-group Multiple Indicators Multiple Causes model (MIMIC). MIMIC is a special case of Structural Equation Modeling (SEM) that tests the impact of covariates on a single-factor measurement model. In the multi-group analysis, we split the data according to the

Mayan-Urban, Mayan-Rural, Ladino-Urban and Ladino-Rural divide (for the mean score in each of the measures by group see Table 2). We also estimated Modification Indexes (MI), indicators of misspecification that represent an expected drop in overall χ^2 value if parameters are freely estimated (Byrne, 2001). We could not pool third and sixth grade data because reading and math tests had no shared items and, consequently, could not be anchored across grade levels.

Table 2. Mean, Standard Deviation and Standardized Mean of Math and Reading Scores per Grade, Area, and Ethnicity

Subject		Third Grade				Sixth Grade			
		Maya		Ladino		Maya		Ladino	
		Urban (n = 2254)	Rural (n = 4751)	Urban (n = 6171)	Rural (n = 7455)	Urban (n = 2088)	Rural (n = 2718)	Urban (n = 5518)	Rural (n = 4903)
Math	M	0.263	-0.241	0.322	0.021	-0.861	-1.294	-0.740	-0.987
	SD	0.903	1.169	0.934	1.034	0.725	0.879	0.718	0.765
	SM	102	97	102	100	101	96	103	99
Reading	M	-0.751	-1.171	-0.647	-0.877	-0.210	-0.699	-0.011	-0.283
	SD	0.737	0.189	0.732	0.782	.688	0.747	0.704	0.708
	SM	101	96	103	100	101	94	103	100
Span. Fam.	M	0.998	-0.668	2.553	2.360	1.07	-0.476	3.507	3.552
	SD	2.400	2.520	1.564	1.814	2.59	2.92	1.128	1.208
	SM	98	91	104	103	94	88	104	104
SES	M	0.128	-0.195	0.131	-0.007	.002	-0.647	0.378	-0.51
	SD	0.930	0.985	0.953	1.023	.936	1.068	0.788	0.985
	SM	101	98	101	100	100	93	104	99

M = Mean; SD = Standard deviation; SM = Standardized mean (rounded to the nearest integer) with an overall arbitrary mean of 100 and a standard deviation of 10.

As fit indexes for the MIMIC model we employed the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), chi-square goodness of fit (χ^2) and the relative chi-square (χ^2/df) (Byrne, 2001). Our decision criteria for a good fit of RMSEA was a value smaller than 0.05, while reasonable fit was greater than 0.05 and smaller than 0.08 (Byrne, 2001). For TLI a good fit required a value equal to or greater than 0.95 (Byrne, 2001). Goodness of fit for χ^2 was reached when the value was statistically non-significant (Byrne, 2001). There is no consensus on the acceptable ratio for the relative chi-square (χ^2/df) and recommendations range from as low as 2 to as high as 5 (Hooper, Coughlan, & Muller, 2008). We decided to adopt the

recommendation of a score of ≤ 2 as reflection of good fit.

The previous analyses should provide us with information of the measurement model, but, depending on sufficient fit, we also wanted to explore how achievement levels differed across the groups and whether exposure to Spanish and SES contributed to understand such differences. In addition, we included gender in this analysis as this variable is associated with exclusion of girls. We conducted a MANCOVA to explore differences in mean scores of math and reading in relation to gender and ethnicity of pupils and area of residence. We ran the MANCOVA two times, first with ethnicity, area, and gender as independent variables and a second time adding exposure to Spanish and SES as covariates.

Results

Third grade RMSEA and CFI indicators showed an adequate fit for the unconstrained, measurement weights, and structural weights models (Table 3). However, χ^2 and χ^2/df exhibited an adequate fit only for the unconstrained model. We estimated Modification Indexes (MI) and identified the suggested relaxation of the constraints for the proposed model. Large MI of the structural weights model were found; the regression weight between achievement and Spanish exposure would produce small changes for Mayan pupils (MI = 9.69 for Urban Mayas and 0.00 for Rural Mayas) and larger changes for Ladino pupils (MI = 20.07 for Urban Ladinos and 42.34 for Rural Ladinos). However, the fit coefficients were not very different after the changes. Other fit statistics (RMSEA, TLI, and CFI) pointed to a good fit of the structural weights model. Therefore, we decided to accept the structural weights model solution as the most restricted model with an adequate fit.

We observed the same pattern in sixth grade as in the third grade: RMSEA, TLI, and CFI indicators were adequate for the unconstrained, measurement weights, and structural weights models, but the χ^2 and χ^2/df statistics were not (see Table 3). As in the case of the third grade, we estimated MIs and found a suggestion for releasing the regression weight between Spanish exposure and achievement in the Ladino Urban group (MI = 38.32). Again, when the invariance constraint

of this parameter was released, the changes were small and inconsequential. Therefore, we decided to accept the structural weights model.

Once the structural weights solution was accepted, all factor loadings were positive and statistically significant ($p < .01$), even in third grade where the SES – achievement regression was comparatively small. The averages of the standardized regression weights for the four groups are presented in Figure 1 for each grade. In both grades the reading – achievement regression was set to 1.00.

In sixth grade the Spanish exposure – achievement regression was statistically significant for Mayan Urban and Mayan Rural groups, but small and not statistically significant for Ladino Urban and Ladino Rural groups (-.02 and .02 respectively). SES – achievement and math – achievement regressions were statistically significant for all sixth grade groups, as were the Spanish exposure – SES covariates. Thus, the results suggest that exposure to Spanish outweighed the importance of SES to predict achievement in third grade. In sixth grade the opposite was true.

We conducted a MANCOVA to explore differences in mean scores of groups on math and reading in relation to gender and ethnicity of pupils and area of residence. We ran the MANCOVA two times, first with ethnicity, area and gender as independent variables and a second time adding exposure to Spanish and SES as covariates. The results are presented in Table 4.

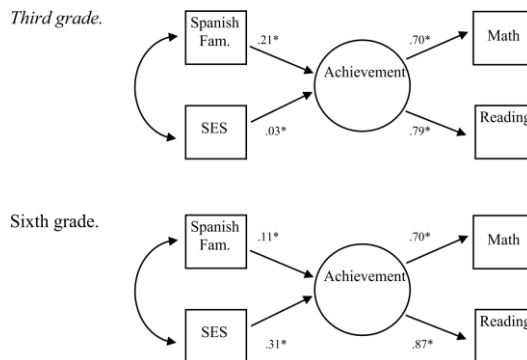


Figure 1. MIMIC model and average standardized weights (structural weights model; achievement – reading fixed at value 1 in non-standardized solution).

* $p < .01$

Table 3. Indicators of fit per grade level for the MIMIC model.

Model	RMSEA	CFI	TLI	χ^2	χ^2/df
Third grade					
Unconstrained	.00	1.00	1.00	4.88	1.22
Measurement weights	.01	1.00	1.00	20.23	2.89
Structural weights	.02	1.00	.99	64.74	5.40
Structural covariances	.07	.81	.81	2065.73	93.90
Structural residuals	.07	.80	.80	2184.19	87.37
Measurement residuals	.06	.78	.78	2413.93	77.87
Sixth grade					
Unconstrained	.00	1.00	1.00	5.08	1.27
Measurement weights	.01	1.00	1.00	19.50	2.78
Structural weights	.02	.99	.99	90.10	6.93
Structural covariances	.13	.37	.38	6011.41	273.25
Structural residuals	.13	.37	.37	6023.05	240.92
Measurement residuals	.12	.35	.35	6247.92	201.55

Table 4. Partial Eta Squared and level of significance for MANCOVA

Effect	Grade	No Covariates	With covariates	$\Delta\eta^2$
		η^2	η^2	
SES	Third		.001***	
	Sixth		.079***	
Spanish Exposure	Third		.031***	
	Sixth		.010***	
Area (A)	Third	.041***	.027***	-.015
	Sixth	.063***	.027***	-.036
Gender (G)	Third	.004***	.004***	.000
	Sixth	.009***	.010***	.001
Ethnicity ϵ	Third	.013***	.002***	-.012
	Sixth	.038***	.002***	-.036
A \times G	Third	.001	.000	.000
	Sixth	.004*	.000*	.000
A \times E	Third	.003***	.001***	-.003
	Sixth	.006***	.002***	-.004
G \times E	Third	.007**	.001***	.000
	Sixth	.002***	.002***	.000
A \times G \times E	Third	.001**	.001***	.000
	Sixth	.000	.000	.000

In third grade our three independent variables (ethnicity, area of residence, and gender) were statistically significant ($p < .001$), both with no covariates added and after these had been included. Interactions were also statistically significant, except the area by gender interaction which was non-significant and small when the covariates had not been added ($F(2, 20631) = 1.22, p = .30$; Wilks' $\Lambda = 1.00$, partial $\eta^2 = .00$), as well as when they had been included ($F(2, 20631) = 1.25, p = .29$; Wilks' $\Lambda = 1.00$, partial $\eta^2 = .00$). Effect sizes (Partial Eta Squared) varied between negligible and small, ranging from .00 to .04 before addition of covariates and from .00 to

.03 with these covariates included (see Table 4). Area of residence exhibited the largest effect before the covariates were added. After covariates were included, area and exposure to Spanish exhibited the largest effect sizes.

In sixth grade, again, all independent variables (ethnicity, area, and gender) were statistically significant at $p < .001$ before and after including the covariates. Interactions were also statistically significant before and after the addition of the covariates, with the exception of the triple interaction of area by gender by ethnicity that was non-significant both before adding the covariates ($F(2, 15227)$

= 0.684, $p = .51$; Wilks' $\Lambda = 1.00$, partial $\eta^2 = .00$) and after including them ($F(2, 15227) = 1.036$, $p = .36$; Wilks' $\Lambda = 1.00$, partial $\eta^2 = .00$) (see Table 4).

Effect sizes (Partial Eta Squared) were negligible to small before adding the covariates (.00 to .04), with one exception; the effect size for area was of medium size (.06). When covariates were added, the effect size for SES was .08 and area dropped to .03, while the remaining effect sizes ranged from .00 to .01. Therefore, area conserved a larger size in comparison to the other independent variables, but still became small in absolute terms.

These results indicate that ethnicity, area, and gender have a statistically significant relationship with the achievement in reading and math. Effect sizes were small, however, and therefore the relevance of these relationships still requires further research, perhaps with tools developed from the onset to measure these particular constructs. This is equally the case for the covariates that also showed statistically significant relationships but small effect sizes (with the exception of sixth grade area before adding covariates and sixth grade SES after adding the covariates). Given the large size of our samples and the small effect sizes, the statistically significant differences we observed become less informative for the design of policies and direct interventions.

The MIMIC model we tested supports the relevance of the relationship of both exposures to Spanish and SES with achievement. It also shows the shift in relevance of these variables from third to sixth grade. While in third grade the differences in exposure to Spanish appear to explain somewhat more variance than SES, in sixth grade the size of these effects is reversed. We found this model to be structurally invariant across groups, which supports the idea that these relationships are consistent across ethnic groups.

Discussion

We conducted this study, in third and a sixth grade samples to examine how linguistic factors, exposure to Spanish in this case, and SES are associated to educational outcomes once the moderating role of ethnicity and area

of residence has been considered. We fitted a MIMIC model where we included measures of exposure to Spanish, SES, reading and math. We ran a multi-group analysis contrasting Mayan Urban, Mayan Rural, Ladino Urban, and Ladino Rural groups of pupils.

We also carried out a MANCOVA for each sample. First we included ethnicity, area of residence (urban/rural), and gender as independent variables, while reading and math performance were the dependent variables. Then we ran a MANCOVA where the independent and dependent variables were the same as before, but we also included exposure to Spanish and SES as covariates. We found the independent variables (area, ethnicity and gender) and the covariates (exposure to Spanish and SES) to hold statistical significance, but the small to medium effect sizes make these results of limited informative value for the design of policy or direct interventions.

We found the relationships in our MIMIC model to be viable and invariant across groups (Mayan Urban, Mayan Rural, Ladino Urban, Ladino Rural), both in third and sixth grade; the relationships between parameters in the model stay consistent. Our fit statistics (RMSEA, TLI, CFI, χ^2 and χ^2/df) were adequate for the structural weights models. Based on this solution (i.e., structural weights), factor loadings were positive (see Figure 1).

The relationships between the latent variable achievement and performance in reading and math tests were strong and positive in third and sixth grade. The regression coefficients for exposure to Spanish and SES showed a somewhat more complex picture. In third grade, the regression between SES and achievement (.03) was small. The regression between exposure to Spanish and achievement was fairly substantial (.21). In sixth grade an inverse pattern was found (the coefficient for exposure to Spanish – achievement was .11 and for SES – achievement was .31). These findings suggest that the relevance of exposure to Spanish decreases as the grade level goes up. We see two possible explanations: the rural Mayan children develop their linguistic ability as they progress through the school system

making the exposure at home and among friends less crucial, or the system differentially selects pupils to continue in the system. While a mixture of both of these explanations could be at play, we believe that in terms of adhering to one for purposes of policy design the main factor underlying the shift is the selectivity of the system. This is compatible with the pattern of change also observed in the national educational statistics that point to a decrease of Mayan Rural pupils (with generally less exposure to Spanish) as they get closer to completing the primary level and lower overall enrolment rates. Therefore, while some factors associated to school quality, such as rate of students per teacher, might improve, the homogeneity of the group also increases providing each pupil with more opportunities to interact in Spanish. By fourth grade all classes are delivered in Spanish, thus decreasing the perceived and real benefit of attending school for those pupils who did not acquire sufficient linguistic ability by third grade. As a result, even if the system is effective in facilitating the acquisition of Spanish linguistic abilities, it is likely that a number of rural Mayan pupils will not have benefitted enough to continue to fourth grade and beyond.

Our MANCOVA findings are consistent with the aforementioned. In both third and sixth grade, area exhibited a statistically significant and moderately sized effect. The largest effect size was found for SES, but only in the sixth grade. In third grade the exposure to Spanish was one of the two variables with a substantial effect size (together with area), while in sixth grade SES showed the larger effect. This also leads us to conclude that area is consistently relevant, but that exposure to Spanish becomes less of a differentiating factor as the grade level progresses. In other words, while area is a very relevant factor in predicting achievement throughout the primary levels, variation linked to the linguistic factor shifts towards SES at higher grades. Again, this is consistent with the shifts in population as grade levels progress at that point to a systematic filtering of pupils who continue education from lower to higher grades, where indigenous children who have less exposure to Spanish are less likely to progress to higher grades.

All in all, we come to two conclusions. First, results confirm what previous research has indicated, namely that SES is a positive predictor of school achievement (e.g., Reynolds, 2003), and that compatibility of the pupils' familiarity with the language of the test has an impact on achievement results (Camilli, 2006). Second, we found that our model suggests a set of relationships that are consistent across groups, whether Mayan or Ladino, Urban or Rural, although mean scores vary. In a country where ethnic differences are consistently highlighted and policy decisions are based on the assumption that there are different psychological mechanisms operating in each group, this introduces a new perspective. The treatment of pupils as homogeneously monolingual or bilingual in accordance with their ethnicity does not lead to a sufficient type of intervention in the heterogeneous Guatemalan population.

In terms of educational expenditure and investment in interventions these results suggest that attacking issues related to SES and the conditions of schools linked to the SES of its pupil population might be the factor to prioritize in policy making. This is not to say that ethnicity is not relevant. It is clear that the ethnicity of the pupil may set differential barriers, particularly as indigenous families tend to be of lower SES, reside in rural areas and have less exposure to Spanish. However, the conditions of the schools seem to be such, that they are not at present equipped to effectively offer adequate services to the children they serve, particularly those more prone to be excluded (pupils of lower SES in rural areas who have less exposure to Spanish). Ensuring that these schools are better equipped, adequately provided with the necessary services, and staffed with teachers who can effectively use the language of the pupils when teaching will improve learning outcomes and increase the likelihood of equitable delivery.

Furthermore, the results provide support for continuing research on linguistic issues in achievement. While there is a three-decade history of bilingual education in Guatemala, there has been little effort to measure actual proficiency in Spanish or the other major languages in a reliable manner. The indication

that exposure to Spanish is associated to school performance and that current policy obliges pedagogical decisions to be devised around pupil's linguistic proficiency, suggests that developing measures of proficiency in Spanish as a second language might be a more viable short-term solution than developing specific measurements for all Mayan languages and their regional variations. We underscore, however, that this is a matter of priorities; we do not deny that adequate development of measures for the other languages is also desirable and that delivery of adequate bilingual education is necessary.

Our study has four main limitations. First, we have considered only two of the variables that compose the complex structure of the context of the pupils (exposure to Spanish and SES) to assess the effects of ethnicity and area of residence. Therefore, our perspective is necessarily limited and further research is required to amplify our vision of the context. However, finding consistency of relationships across groups already provides some orientation to propose designs that target not only differences between groups, but also similarities.

Second, the attribution of ethnicity that we employed is based on pupil self-identification. We were unable to explore how this ascription to a particular ethnicity is developed or how schooling affects it. We are not aware, however, of systematic research having been conducted on this issue within the school context in Guatemala and we propose this as an area for further study.

Third, we had a number of missing cases that required imputation. These amounted to 25% of the math scores in third grade, 34% of the reading scores in third grade, 18% of the math scores in sixth grade and 33% of the reading scores in sixth grade. This requires us to ponder our findings with care. Rates of missing cases in national assessments have gradually improved, but these particular grade levels have not been assessed every year and questionnaires usually have not offered the indicators required to conduct the analyses we presented here. Given these conditions and the difficulty in obtaining data on this subject for Guatemala, we concluded that moving ahead

with the study was justified and that the insights it provides can be a worthwhile contribution for policy development if the necessary caveats are made. The results are of immediate interest for the Guatemalan educational system, but also offer interesting inputs for a wider context. Most countries where pupils are schooled in multi-lingual environments will also have to negotiate between the use of their mother language, the language of the education system, and the different levels of mastery of each language across individuals.

Fourth, a more accurate measure of SES might permit clearer interpretations. For example, the Standardized Mean of SES for rural third graders is 98, higher than we would have anticipated since the Standardized Mean of SES for rural sixth graders is 93. Furthermore, Cronbach's alpha for the third grade scale was 0.56 and 0.57 for sixth grade. The data we employed was the best available data to provide an approximate measure of the SES of these pupils, but evidently a more reliable and valid measure could be developed to gather more informative results on the relationships of SES to achievement.

We offer two recommendations for educational policy based on our findings. First, schools offering bilingual education should have a clear perspective regarding their pupils' language mastery. This does apply to L1 (mother tongue), but also to the fluency in Spanish. Rather than assuming that Spanish is being acquired through the first three grades of schooling, teachers should be aware of the level of Spanish that pupils have at the onset and whether the level in L1 is comparable to the level in Spanish. Second, although the bilingual-ethnic divide is an important one, the rural-urban split should continue to be explored and stressed and SES conditions prioritized. There is little understanding of the differential demands of rural and urban settings and how these interact with the demands of formal schooling and school quality in bilingual contexts, while it is clear that low SES has a detrimental effect on the schooling of pupils. Further research is required on what exactly are the implications of living and studying in rural areas and the SES status in the cultural and educational

practices of Mayan and Ladino pupils as they interact in the linguistically heterogeneous setting of formal schooling.

References

- Álvarez, H., & Schiefelbein, E. (2007, December). *Informe integrado del sector educación: Informe final [Consolidated report of the education sector: Final report]*. (Report financed by the Interamerican Development Bank and the Swedish Agency for International Development as input for the strategy on Policy Harmonization and Alignment). Guatemala: Ministry of Education (BID / MINEDUC / ASDI).
- Anuario Estadístico de Educación 2006 [Annual Educational Report 2006]*. (n.d.). Retrieved from <http://estadistica.mineduc.gob.gt/anuario/2006/home.htm>
- Beckett, M., & Pebley, A. R. (2003). Ethnicity, language and economic well-being in rural Guatemala. *Rural Sociology*, 68, 434–458. doi:10.1111/j.1549-0831.2003.tb00145.x
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. Mahwah, NJ: Erlbaum.
- Bond, T. G., & Fox, Ch. M. (2007). *Applying the Rasch model* (2nd ed.). Mahwah, NJ: Erlbaum.
- Camilli, G. (2006). Test fairness. In Brennan, R. L. (Ed.), *Educational Measurement* (4th ed.). (pp. 221–256). Westport, CT: American Council on Education and Praeger Publishers.
- Chesterfield, R. A., Enge, K. I., & Rubio, F. E. (1997). Cross-cultural cognitive categorization of pupils by Guatemalan teachers. *Cross-Cultural Research*, 36, 103–122. doi:10.1177/1069397102036002001
- Correa-Chávez, M., & Rogoff, B. (2009). Children's attention to interactions directed to others: Guatemalan Mayan and European American patterns. *Developmental Psychology*, 45, 630–641. doi:10.1037/a0014144
- Enge, K. I., & Chesterfield, R. (1996). Bilingual education and pupil performance in Guatemala. *International Journal of Educational Development*, 16, 291–302. doi:10.1016/0738-0593(95)00038-0
- Esquivel Villegas, F. (2006, February). *Situación del sistema educativo guatemalteco [Condition of Guatemalan educational system]*. (Consultancy document developed for the Ministry of Education and the World Bank in preparation of support documents for the discussion on a new loan). Guatemala, Guatemala: Ministry of Education of Guatemala and National Office for Guatemala of the International Bank for Reconstruction and Development.
- Flores Reyes, M. G., & Villegas de Sanders, A. R. (2009). Estudio etnográfico: Oportunidades de aprendizaje del idioma materno [Ethnographic study: Opportunities to learn the mother language]. Retrieved from http://www.mineduc.edu.gt/recursoseducativos/descarga/digeduca/investigaciones/ficha_5.pdf
- Fortin Morales, A. M., van de Vijver, F. J. R., & Poortinga, Y. H. (2013). Differential item functioning and educational risk factors in Guatemalan reading assessment. *Interamerican Journal of Psychology*, 47, 423–432.
- Fuller, B., & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy. *Review of Educational Research*, 64, 119–157. doi:10.3102/00346543064001119
- Fuller, B. (1987). What school factors raise achievement in the Third World? *Review of Educational Research*, 57, 255–292.
- Hambleton, R. K., Merenda, P. F., & Spielberger, Ch. D. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Erlbaum.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6, 53–60. Retrieved from <http://arrow.dit.ie/cgi/viewcontent.cgi?article=1001&context=buschmanart>
- INE (2010). *Perfil estadísticos de género y pueblos: maya, garífuna, xinka y ladino [Statistical profiles of gender and peoples: Maya, Garífuna, Xinka and Ladino]*. Retrieved from <http://www.ine.gob.gt/np/generoypueblos/documentos/Perfil%20Estadistico%20final.pdf>
- Linacre, J. M. (2008). WINSTEPS (Version 3.66) [Computer software]. Chicago, IL: Winsteps.com
- Marzano, R. J. (2006). *Classroom assessment & grading that work*. Alexandria, Alexandria, VA: Association for Supervision and Curriculum Development.
- MINEDUC/DIGEDUCA (n.d.). *Informe Primaria 2006 [2006 Report for Primary Level]*. (Report of the national assessment of a sample of primary level official/public schools of Guatemala conducted by the Directorate for Assessment of Educational Quality of the Ministry of Education of Guatemala). Guatemala, Guatemala: Ministry of Education. Retrieved <http://www.mineduc.gob.gt/digeduca/>
- Moreno Grajeda, M. R. (2009). GINI Educativo en Guatemala, en sus departamentos y su relación con variables de desarrollo [Educational Gini in Guatemala, in its Departments and its relationship with human development variables]. Retrieved from http://www.mineduc.edu.gt/recursoseducativos/descarga/digeduca/investigaciones/ficha_6.pdf
- Moreno Grajeda, M. R. (2012). La desigualdad social y el aprendizaje de lectura y matemáticas en estudiantes de sexto grado de primaria [Social inequality and reading and math achievement in sixth grade pupils]. Retrieved from http://www.mineduc.gob.gt/digeduca/documents/investigaciones/DESIGUALDAD_SOCIAL_Y_APRENDIZAJE.pdf
- Patrinos, H. A., & Velez, E. (2009). Costs and benefits of bilingual education in Guatemala: A partial analysis. *International Journal of Educational Development*, 29, 594–598. doi:10.1016/j.ijedudev.2009.02.001
- Pellecer Rivera, H. S., Saz Choxin, M. A., & Santos Solares, J. A. (2011). Desigualdad en la cobertura educativa en Guatemala medida a través del coeficiente de GINI y la Curva de Lorenz [Inequality in educational coverage in Guatemala measured through the GINI coefficient and the Lorenz Curve]. Retrieved from <http://www.mineduc.gob.gt/digeduca/documents/inv>

- estigaciones/Informe_Gini_Educativo_2009_Terminado.pdf
- Población en Guatemala (demografía). (n.d.). In *INE (Instituto Nacional de Estadística [National Institute for Statistics])*. Retrieved from <http://www.ine.gov.gt/np/poblacion/>
- Poortinga, Y. H. (1989). Equivalence of cross-cultural data: An overview of basic issues. *International Journal of Psychology*, 24, 737–756. doi:10.1080/00207598908247842
- Reynolds, D. (2003). School effectiveness: The international dimension. In Ch. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 231–256). London, United Kingdom: Falmer Press.
- Rubio, F. E. (2004). *Educación bilingüe en Guatemala: Situación y desafíos [Bilingual education in Guatemala: Situation and challenges]*. Paper presented for the workshop / seminar Balance y Perspectivas de la Educación Intercultural Bilingüe [Balance and Perspectives from Intercultural Bilingual Education] of the World Bank-PROEIB Andes, Guatemala, Guatemala. Retrieved from <http://psetecnologiaplicada2011.files.wordpress.com/2013/04/educacion-bilingue-en-guatemala.pdf>
- Sireci, S. G. (2011). Evaluating test and survey items for bias across languages and cultures. In Matsumoto, D., & van de Vijver, F. J. R. (Eds.), *Cross-cultural research methods in psychology* (pp. 216–243). Cambridge, United Kingdom: Cambridge University Press.
- Sistema Nacional de Indicadores Educativos [National System of Educational Indicators]*. (n.d.). Retrieved from <http://estadistica.mineduc.gov.gt>
- Treviño Villareal, E. (2006). Evaluación del aprendizaje de los estudiantes indígenas en América Latina: Desafíos de medición e interpretación en contextos de diversidad cultural y desigualdad social [Evaluation of educational achievement of indigenous pupils in Latin America: Challenges of measurement and interpretation in culturally diverse and socially inequitable contexts]. *Revista Mexicana de Investigación Educativa*, 11, 225–268. Retrieved from <http://www.redalyc.org/articulo.oa?id=14002812>
- UNDP (2014). Statistical Table for the 2014 Human Development Report. [Online interactive data-base.] Retrieved from <https://data.undp.org/dataset/Table-4-Gender-Inequality-Index/ku9i-8fxp>
- van de Vijver, F. J. R. (2011). Bias and real differences in cross-cultural differences: neither friends nor foes. In F. J. R. van de Vijver, A. Chasiotis, & S. M. Breugelmans (Eds.), *Fundamental questions in cross-cultural psychology* (pp. 235–257). Cambridge, United Kingdom: Cambridge University Press.
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée*, 54, 119–135. doi:10.1016/j.erap.2003.12.004
- World factbook*. (2007). USA: CIA. Retrieved from <https://www.cia.gov/library/publications/the-world-factbook>