Intelligence 98 (2023) 101758

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Bilingualism and intelligence in children exposed to poverty environments: A Raven's error pattern analysis using a generalized propensity score method

Leopoldo Laborda[a,*], Juan Mejalenko[b], Isabel Gómez-Veiga[c]

[a] Universidad Nacional de Educación a Distancia (UNED), Escuela Internacional de Doctorado (EIDUNED), C/ Bravo Murillo, 38, 28015 Madrid, Spain
[b] The University of Chicago Booth School of Business, 5807 S Woodlawn Ave, Chicago, IL 60637, United States
[c] Universidad Nacional de Educación a Distancia (UNED), Facultad de Psicología, C/ Juan del Rosal, 10, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Intelligence
Bilingualism
Poverty
Raven's error patterns
Generalized propensity score (GPS) method

ABSTRACT

The main objective of this study is to compare the quantitative (correct answers) and qualitative (error types) performance of children belonging to different linguistic groups on a non-verbal reasoning test, Raven's Colored Progressive Matrices Test, after being matched based on level of exposure to poverty, certain individual characteristics and test performance. The sample is representative for Mexico at a population level and consists of children aged 5 to 12 ($n = 4644$), of which 671 are bilingual in Spanish and an indigenous language, 3970 are monolingual in Spanish and 78 are monolingual in an indigenous language. The results show significant quantitative differences with a lower overall performance in the Raven's test by bilingual children as compared to their monolingual (Spanish only) peers, but no qualitative differences when analyzing their error types. When considering each linguistic group individually, the relative frequency of three error types (Repetition, Wrong Principle, and Incomplete Correlate) is similar in children aged 5 to 8 and in those aged 9 to 12. However, considering the two age cohorts, the results reveal how the intragroup differences in each linguistic group, are only statistically significant in the case of Difference errors, in the group of monolingual children in Spanish.

In addition to practical use that may be potentially derived from this empirical evidence, these results may also be encouraging from a methodological point of view. They demonstrate how the method used, in addition to permitting greater comparison between the experimental groups of a representative sample at a population level, does not present high sensitivity, either for the model used to estimate the Generalized Propensity Score method, or for the specific estimator used.

1. Introduction

Individual differences in intelligence are often measured using psychometric tests, such as the Raven's Standard Progressive Matrices Test (SPMT: Raven, Court, & Raven, 2006). The SPMT uses visuospatial information processing to solve visual analogy problems of increasing difficulty. The Raven's inductive reasoning task is assumed to measure fluid intelligence (g_f), defined as the ability to solve novel reasoning problems independent of knowledge from the past (Cattell, 1963). A priori, this test is less influenced by, although not exempt from, cultural differences as compared to other tests in which language is directly involved (Gonthier, 2022). Nowadays, a consistent finding is that children's performance on intelligence tests, such as SPMT, is influenced by a variety of factors, specifically, genetics and maturation (Deary, Penke, & Johnson, 2010; Haier, 2016). Evidence also suggests that environmental factors may affect cognitive functioning and development (e.g., Crosnoe et al., 2010). This article explores the role of childhood exposure to poverty and the regular use of two languages (bilingualism, in this paper) on cognitive performance in a test assumed to measure fluid aspects of intelligence, such as the Raven's Colored Progressive Matrices Test (CPMT: Raven et al., 2006). Although poverty has been commonly examined as a risk factor for cognitive development (Lipina, 2016), bilingualism has been proposed as a potential cognitive advantage (Kroll & Bialystok, 2013a, 2013b). We examined the relationships between the variables in a wide sample of children growing up in conditions of poverty in Mexico.

* Corresponding author.
E-mail addresses: llaborda1@alumno.uned.es, llabordacastillo@gmail.com (L. Laborda), igveiga@psi.uned.es (I. Gómez-Veiga).

https://doi.org/10.1016/j.intell.2023.101758
Received 3 October 2022; Received in revised form 12 April 2023; Accepted 12 April 2023
Available online 25 April 2023

Childhood poverty is a complex construct involving individual, family, and environmental factors that are associated with material, emotional and symbolic deprivation (Lipina, 2016; Lipina & Evers, 2017). Of the environmental factors, research on this topic tends to examine the role of the family's economic well-being (i.e., income level, main source of family income, health care, household conditions, and parental education), as well as the quality of the school environments, early childhood program attendance, and health and nutrition on cognitive development. Past studies have identified associations between lower scores on measures of economic well-being and lower scores on cognitive development. It has been suggested that poverty can influence cognitive performance during childhood in a number of ways. Specifically, a systematic review by Segretin et al. (2016) examined 53 studies investigating the relationship between poverty and cognitive development in children under 18, identifying a pattern of the relationship between poverty and low performance on children's cognitive measures, including Raven's test. More recently, Platt et al. (2018), using a wide sample of adolescents aged 13–18, examined the relationship between fluid intelligence test performance and deprivation experiences related to access to adequate social services and material conditions —including poverty and low parental education—. The results supported associations between the Socio-Economic Status (SES) and cognitive ability in children, including fluid intelligence (Lynn, Fuerst, & Kirkegaard, 2018). The authors concluded that social, perceptual, and linguistic stimulation, as well as opportunities for varied activities, are reduced in children raised in low-SES environments, and that limitations in these stimulations may adversely influence cognitive development.

These findings are consistent with those of past studies that have shown that poverty at any developmental stage is associated with lower performance on intelligence tests (e.g., Anum, 2022; Ayoub et al., 2009; Kaya, Stough, & Juntune, 2016), even though the impact of childhood poverty on cognition depends on the timing, sequence, and duration of exposure to deprivations (Lipina, 2016). In that respect, Ryan, Fauth, and Brooks-Gunn (2006) estimated that children who come from poor households answer 15–40% fewer Raven's questions correctly than those of their peers from non-poor households. In a longitudinal study, Najman et al. (2009) concluded that prolonged exposure to an impoverished environment from childhood to adolescence would be more detrimental to cognitive outcomes than when occurring over shorter periods, estimating that, for each additional period of exposure to poverty, the overall scores on the Raven's SPMT would be reduced by 2.19 points. The relationship between poverty and Gf was found to exist, even when controlling for the effect of maternal cognitive skills (Gredebäck, Hall, & Lindskog, 2022; Wachs, Chang, Walker, & Gardner, 2007), parental educational level (Rahu, Rahu, Pullmann, & Allik, 2010) and family structure (Tong, Baghurst, Vimpani, & McMichael, 2007). More recently, Daniele (2021) reported evidence showing how poverty rates and educational inequality in income are significantly and negatively related to educational outcomes (i.e., PISA scores). Currently, it is important to consider that, as Segretin et al. (2016) suggested, the effect of poverty may be modulated and moderated by systematic differences among individual and societal and cultural patterns of parenting, schooling, and psychosocial environment, as well as variations in the risk factors to which the children are exposed, the timing of exposure to risk factors, and the individual susceptibility to each variable. In other words, poverty does not necessarily generate homogeneous and continuous changes in all measures of neurocognitive processing and this is not uniform across all ages (Lipina, 2016).

This work has considered the effects of poverty on specific populations, and the results of numerous studies carried out on the performance of indigenous children in non-verbal reasoning tests (Millones, Flores-Mendoza, & Rivalles, 2015; Nistal, 2014). However, given that in previous studies, ethnic condition has been associated with a level of poverty, this raises the question of whether these results may be replicated once greater equivalence is established between the indigenous and non-indigenous groups, matching not only certain variables of material poverty but also considering other factors related to the family environment in which the children grow up. Therefore, it has recently been verified that, when equating samples of poor indigenous and non-indigenous children in terms of age, parental education level, assistance received in performing school tasks, and some specific parenting patterns, intergroup differences in non-verbal intelligence measured with the Raven's Progressive Colored Matrices Test (CPMT) are reduced after matching (Laborda, Elosúa, & Gómez-Veiga, 2019).

As for the second variable mentioned above, bilingualism has been defined as achieving a state of communicative knowledge of two or more languages (Grosjean & Li, 2013). In addition to communication advantages, bilingualism has been associated with certain cognitive advantages, specifically, executive functioning, resulting from the need of the bilingual individual to monitor his/her circumstances in order to select the suitable target language (Bialystok, Craik, Green, & Gollan, 2009). Therefore, extending beyond language processing to other aspects of cognitive functioning, the possible benefits of bilingualism have been mainly highlighted regarding the role of the executive function (for a review, see, Bialystok, 2017; Lowe, Cho, Goldsmith, & Morton, 2021).

Specifically, numerous studies comparing monolinguals and bilinguals have revealed better performance by bilinguals on tasks that appear to measure inhibitory control, selective attention, cognitive flexibility, working memory, and even problem-solving (Adesope, Lavin, Thompson, & Ungerleider, 2010). If this is so, the linguistic and cognitive results of the regular use of two languages may be intimately interconnected and involve the reorganization of complex mental structures in response to a specific linguistic experience (Kroll & Bialystok, 2013a, 2013b). In fact, neuroimaging results have revealed that the neural processing of bilinguals and monolinguals differs during the performance of executive functioning tasks (Paap, 2019), suggesting that the coordination of two languages leads to a reorganization of neural networks involved in language control and executive functions, but not implying more efficient behavioral performance.

The evidence, however, is unclear and an open debate exists regarding the so-called *bilingual advantage* (Kroll & Bialystok, 2013a, 2013b). While the results of some studies support the positive effects of bilingualism on cognition, others have failed to find consistent evidence for this advantage (for a review, see, Antoniou, 2019; Bialystok, 2017; De Bruin, Dick, & Carreiras, 2021; Van den Noort, Struys, & Bosch, 2019). Therefore, enhanced cognitive functioning in bilinguals has been recently questioned, since a number of meta-analyses have indicated that the cognitive bilingual advantage is small or may only exist in specific circumstances or for specific types of bilinguals (Antoniou, 2019; Donnelly, Brooks, & Homer, 2019; Gunnerud, Ten Braak, Reikerås, Donolato, & Melby-Lervåg, 2020; Lehtonen et al., 2018; Paap, 2019). Across studies, bilingual and monolingual groups often vary in factors such as literacy in each language, ethnicity, and cultural/social background. In addition, tasks often vary across studies. Therefore, authors such as Bialystok et al. (2009) note that multiple factors make the bilingual experience heterogeneous. This means that further analyses of the social, cognitive, and personal factors that may influence language experiences underlying bilingual-monolingual differences are required. We are especially interested in analyzing how the regular use of two languages may reveal differences in Raven's PMT performance between bilingual and monolingual children growing up in an impoverished environment.

Existing evidence on the joint effects of bilingualism and poverty on cognitive performance in intelligence tests such as the Raven's SPMT date back to the pioneering work of Ben-Zeev (1977a, 1977b) with children aged between 4 and 8 years old, from different language groups in Israel and the United States. In the first of these studies, middle-class bilingual Hebrew-English speaking children were compared with their monolingual peers, revealing similar performance by both groups (Ben-Zeev, 1977a). In the second study, however, living in a disadvantaged urban neighborhood was also considered a factor and, in this case, it was observed that bilingual Spanish-English speaking children from disadvantaged neighborhoods had lower performance than their monolingual peers, albeit with similar but attenuated error patterns compared to their peers from non-disadvantaged neighborhoods (Ben-Zeev, 1977b). More recently, Engel de Abreu, Cruz-Santos, Tourinho, Martin, and Bialystok (2012) found no significant differences in the overall performance on the Raven's SPMT administered to bilingual Portuguese-Luxembourgish-speaking children from low-income immigrant families from Luxembourg and their monolingual peers from Portugal.

Authors such as Kaushanskaya and Prior (2015), however, warned that the approach adopted in some studies (e.g., Engel de Abreu et al., 2012) of matching the socioeconomic level of the participants does not eliminate the problem of comparing groups that may differ in other variables (e.g., general language skills, family size, etc.) influencing executive function development and, ultimately, cognitive performance on tests fluid intelligence tests. These methodological challenges, linked to a correct comparison of experimental and control groups regarding linguistic and socioeconomic characteristics, are not the only ones that have been mentioned in the literature. Thus, it is worth noting that, recently, authors such as Bialystok and Shorbagi (2021) have proposed another approach focusing on the analysis of subtle differences in levels of bilingualism and exposure to poverty as an alternative strategy to the more traditional approach of comparing clearly disparate groups. In the aforementioned study, the results obtained by 6-year-old Canadian children on the Raven's SPMT did not reveal significant effects based on socioeconomic level, level of bilingualism, or the interaction of both.

In contrast to the extensive literature on the overall Raven's task performance, as a proxy and quantification of non-verbal abstract reasoning ability, disaggregated analyses of the error patterns during problem solution are less available, especially for bilingual children exposed to poverty. It is assumed that erroneous responses could be useful indicators of the process and strategies used by children when completing tasks, and may be useful for revealing differences between groups (Raven et al., 2006). Raven and collaborators identified four main types of erroneous responses: 'Difference' errors (D), in which the answer was a piece that either has no pattern or is incongruent with the target matrix; 'Repetition' errors (R), occurring when the answer reproduces a figure immediately above or beside the gap in the matrix; 'Wrong Principle' errors (WP), in which the chosen element is inconsistent, incongruent or incomplete for the target matrix to be completed; and 'Incomplete Correlate' errors (IC), indicating an incorrectly orientated or incomplete but correct figure. Since the work by Ben-Zeev (1977b) analyzing the error patterns of bilingual children from disadvantaged neighborhoods, subsequent evidence has focused mainly on the study of different populations that consist of children of more typical development (Farran, Atkinson, & Broadbent, 2016; Gunn & Jarrold, 2004; Kunda, Soulières, Rozga, & Goel, 2016; Petretto, Grassi, Masala, & Nicotra, 2021; Van Herwegen, Farran, & Annaz, 2011) and individuals with a developmental disorder, such as Down Syndrome (Gunn & Jarrold, 2004), Williams Syndrome (Farran et al., 2016; Van Herwegen et al., 2011), Autism Spectrum Disorder (Kunda et al., 2016), and some other type of Intellectual Disability (Facon, Magis, Nuchadee, & De Boeck, 2011; Goharpey, Crewther, & Crewther, 2013), and with Moderate Learning Difficulties (Gunn & Jarrold, 2004) (see Table 1, for a synthesis of the main results of previous studies). Specifically, Matzen, Van der Molen, and Dudink (1994) reported that typically developing children ($n = 1655$, aged between 8.5 and 12.5 years) made, regardless of the level of ability, a high proportion of IC errors, followed by WP, R, and errors named as additional elements (similar to D errors). However, children with high and low performance on the Raven's test differed in IC and R errors. Moreover, Gunn and Jarrold (2004) found a significant change with age in the types of errors produced by children with typical development so that, as age increases, the proportion of D and WP errors decreases while the proportion of R and IC errors increases. Later, these results were partially validated by Van Herwegen et al. (2011), who obtained similar outcomes, except for IC errors, which did not have a statistically significant relationship. Recently, Petretto et al. (2021) analyzing children having typical developmental patterns ($n = 780$), found that older children (aged 6.5 to 7.5 years) had a higher overall Raven's performance than younger ones (aged 5.5–6.5 years), with R errors being more frequent when considering both the younger group and the total sample (Gunn & Jarrold, 2004; Van Herwegen et al., 2011).

For our purposes, the study by Ben-Zeev (1977b) on bilingual children in poverty conditions is of special interest. It is currently known that, although bilingual children in poverty conditions may display similar performance on overall measures of the Raven's SPMT as compared to their poor monolingual peers, both groups present differences in certain error patterns when completing this test. Specifically, poor bilingual children tend to make fewer errors than their monolingual peers on the so-called "scan" failures. The argument presented in the aforementioned study to explain this type of error claims that bilingual children approach the task as an ill-defined problem with a resolution strategy or, at least, with greater attention to the possibility of differences and to the nature of the differences. These results were observed, despite deficiencies in the use of vocabulary and syntax by the Spanish-English bilingual children as compared to the control group of similar ethnic and social origin.

The lack of empirical evidence regarding the error patterns of poor bilingual children on the Raven's SPMT contrasts with the fact that poverty and bilingualism are often associated with belonging to certain minority groups in many countries. This phenomenon represents a non-negligible percentage of the total population of countries such as Mexico, from which we have extracted the sample for this study. In Mexico, according to the National Institute of Statistics and Geography (INEGI, 2021), 21.5% of the population (25.7 million people) define themselves as indigenous, while 6.5% of the national population (7.4 million people) speaks an indigenous language. On the other hand, indigenous communities continue to be the most vulnerable in terms of inequality. According to the National Council for the Evaluation of Social Development Policy (*Consejo Nacional de Evaluación de la Política de Desarrollo Social* - UNICEF-CONEVAL, 2019), 69.5% of the indigenous population (8.4 million people) experience a situation of poverty and 27.9% (3.4 million people) live in extreme poverty. The Mexican indigenous population has less access to basic healthcare services than the non-indigenous population (Gutiérrez et al., 2020). Another interesting fact is that 43% of speakers of an indigenous language have not completed primary education (Oxfam, 2018).

The main objective of this study is to analyze the overall performance and pattern of errors on the Raven's CPMT of bilingual children exposed to impoverished environments, and to determine whether their performance differs from that of their monolingual peers (Spanish and indigenous language speakers), who have been matched with respect to the level of exposure to poverty, the type of household area, and individual characteristics as sex and age.

Therefore, this study could provide valuable information not only on the possible quantitative differences (overall performance) between bilingual and monolingual groups (Gunn & Jarrold, 2004), but also on potential qualitative differences (types of errors made when responding incorrectly and the frequency of each type) derived from the processing of the information that underlies the different strategies used to solve this type of reasoning tasks (Kunda et al., 2016; Kunda, Soulières, Rozga, & Goel, 2013). To carry out this study, the database provided by the National Survey on Household Living Standards in Mexico (*Encuesta Nacional sobre Niveles de Vida de los Hogares en México* - MxFLS) was used. In addition to being representative at a population level, this database also has a relatively large sample size to analyze the existence of statistically significant differences at all levels of comparison between the different groups of children aged between 5 and 12 years. Taking advantage of the considerable amount of data provided by the MxFLS on household characteristics and the activities carried out by their members, we have constructed a multidimensional poverty measure that includes factors associated with both material and symbolic poverty.

Table 1
Previous studies and main results on error types of children in Raven's Progressive Matrices Test.

Study	Sample	Age in years: mean (SD)	Country	Full scale IQ	PM Raven Score	Proportion of different conceptual types			
						Repetition (R)	Difference (D)	Wrong Principle (WP)	Incomplete Correlate (IC)
Petretto et al. (2021)	TD = 183 TD = 203 All = 386	5.5–6.5 6.5–7.5	Italy		CPM = 18.72 (4.40) CPM = 20.27 (4.74) CPM = 19.58 (4.64)	0.2867 0.2630	0.0450 0.0383	0.0823 0.0901	0.00823 0.0753
Kunda et al. (2016)	TD = 54 ASD = 105	11.96 (3.40) 11.02 (2.99)	Canada	109.82 (10.35) 84.38 (20.03)	CPM = 42.61 (9.79) CPM = 37.57 (12.13)	0.005 & 0.01 0.02 & 0.025	0.015 & 0.02 0.01 & 0.015	−0.02 (ns) −0.02 (ns)	−0.005 & -0.01 −0.01 & -0.015
Farran et al. (2016)	WS = 24 TD = 20 TD = 18 TD = 18 TD = 19	20;5[1] (5:03) 4;08 (0;02) 5;07 (0;03) 6;05 (0;04) 7;06 (0;04)	UK		CPM = 18.92 (4.66) CPM = 11.20 (4.81) CPM = 16.06 (5.41) CPM = 19.61 (4.82) CPM = 5.13 (19.00)				
Kunda et al. (2013)	TD = 54 ASD = 108 ASTI = 96 Matched TD = 38 ASD = 38 ASTI = 38	11.96 (3.40) 11.02 (2.99) n.a 11.11 (3.30) 10.76 (2.71) n.a	Canada	109.82 (10.35) 84.38 (20.03) n.a 106.08 (9.08) 88.83 (18.79) n.a	CPM = 38.26 (8.07) CPM = 38.26 (8.09) CPM = 38.29 (8.07)	0.2 & 0.3 (ns) 0.2 & 0.3 (ns) 0.3 & 0.4	0.1 & 0.2 (ns) 0.1 & 0.2 (ns) 0.0 & 0.1	0.3 & 0.4 (ns) 0.3 & 0.4 (ns) 0.4 & 0.5 (ns)	0.1 & 0.2 (ns) 0.1 & 0.2 (ns) 0.1 & 0.2 (ns)
Facon et al. (2011)	ID = 460 TD = 488	168.08[2] (44.13) 67.10 (9.34)	France		CPM = 16.57 (5.42) CPM = 18.59 (5.32)				
Van Herwegen et al. (2011)	WS = 53 TD = 53	18;03[1] (9.10) 5;8 (1;03)	UK		CPM = 19.34 (4.91) CPM = 19.68 (5.08)	0.5 & 0.6 (ns) 0.5 & 0.6 (ns)	0 & 0.1 (ns) 0 & 0.1 (ns)	0.1 & 0.2 (ns) 0.1 & 0.2 (ns)	0.1 & 0.2 (ns) 0.1 & 0.2 (ns)
Fajgelj, Bala, & Katić (2010)	TD = 116 TD = 341 TD = 512 TD = 421 TD = 229 TD = 249 TD = 250 TD = 216	4 5 6 7 8 9 10 11	Serbia		CPM = 14.88 (3.61) CPM = 18.28 (4.57) CPM = 20.79 (5.75) CPM = 24.38 (5.79) CPM = 27.33 (5.53) CPM = 27.94 (5.89) CPM = 29.30 (5.41) CPM = 31.27 (4.52)				
Facon and Nuchadee (2010)	TD = 48 DS = 48 ID = 48	5.40 (14.46) 17.18 (36.61) 17.01 (31.67)	France		CPM = 13.19 (4.78) CPM = 13.19 (4.78) CPM = 13.19 (4.78)				
Najman et al. (2009)	NP = 1486 P1 = 713 P2 = 427 P3–4 = 311	14 14 14 14	Australia		PMS = 102.33[a] PMS = 100.09 [b] PMS = 98.30 [c] PMS = 95.52 [d]				
Gunn and Jarrold (2004)	DS = 39 MLD = 171 TD = 213 Matched DS = 39	13;0[1] (29.10) 11;10 (33.30) 7;5 (18.01) (months) 156.03 (29.3)	UK		CPM = 13.15 (3.32) CPM = 13.21 (2.54)	0.4 & 0.6 0.6 & 0.8 0.6 & 0.8 0.4 & 0.6 0.4 & 0.6 0.4 & 0.6	0.0 & 0.2 0.0 & 0.2 0.0 & 0.2 0.0 & 0.2 0.0 & 0.2 0.0 & 0.2	0.2 & 0.4 0.0 & 0.2 0.0 & 0.2 0.2 & 0.4 0.2 0.2	0.2 0.2 0.2 0.2 0.2 0.2

(continued on next page)

Table 1 (*continued*)

Study	Sample	Age in years: mean (SD)	Country	Full scale IQ	PM Raven Score	Proportion of different conceptual types			
						Repetition (R)	Difference (D)	Wrong Principle (WP)	Incomplete Correlate (IC)
	MLD = 62	123.96 (31.81)			CPM = 13.20 (1.86)				
	TD = 50	71.31 (10.68)							
Kaniel and Fisherman (1991)	*I* = 250	14.7 (1.67)	Israel		PMS = 27 (5–10 P)				
	NI = 165	15.5 (0.14)			PMS = 45 (50 P)				
	NI = 137	13.35 (0.15)			PMS = 43 (50 P)				
	NI = 361	12.75 (0.15)			PMS = 39 (40 P)				
	NI = 304	11.25 (0.13)			PMS = 35 (45 P)				
	NI = 363	10.5 (0.26)			PMS = 32 (50 P)				
	NI = 410	9.65 (0.15)			PMS = 28 (50 P)				
Jacobs Paul & Mary (1970)	NJ = 45	First grade	USA		CPM = 15.8				
	NJ = 36	Second grade	Baffin		CPM = 18.1				
	NJ = 20	Third grade	Island		CPM = 21.5				
	E = 114	10 and over[3]	Sierra		CPM = 27.0				
	T = 119	10 and over[3]	Leone		CPM = 13.3				

Notes: TD = Typically Developing; ASD = Autism Spectrum Disorder; WS = Williams Syndrome; DS = Down Syndrome; MLD = Moderate Learning Disabilities; ASTI = Computational model (Kunda et al., 2013); CPM = Raven's Colored Matrices test; SPM = Standard Progressive Matrices; ID = Intellectual Disability. NID = No-Intellectual Disability. I = Immigrant; NI = Non-Immigrant. P = Percentile; E = Eskimo. T = Temne. 10 and over from five age groups (10–15; 16–20; 21–30; 31–40; and over 40. G(H) = Gentofte is an SES above the Danish average; E(A) = Esbjerg is about the average of the Danish SES. NP = Never poverty; P1 = poverty 1 period; P2 = poverty 2 periods; P3 = poverty 3–4 periods; [a] [74.7–129.9 95% C·I]; [b] [71.0–129.2 95% C·I]; [c] [67.9–128.7 95% C·I]; [d] [63.7–127.3 95% C·I].
Source: Prepared by the authors.

A novel methodological strategy for this field was used in this study, based on the comparison of bilingual participants with their monolingual peers, according to individual (i.e., sex and age) and socioeconomic and demographic variables. Specifically, the Generalized Propensity Score (GPS) method was applied. This method is especially appropriate for conducting an analysis such as the one proposed here. As will be discussed later, this approach permits the combining of the "Propensity-Score Matching" (PSM) method with multiple treatments (bilingual/Spanish monolingual/Indigenous language monolingual) when it is impossible to know their level of intensity and/or impose conditions of ignorability (Flores & Mitnik, 2013). Thus, we can minimize the bias between the different groups resulting from exposure to poverty and their demographic and individual characteristics (sex and age). In addition, to analyze the error patterns of the participants when completing the Raven's CPMT and to avoid biased interpretations, this methodological strategy allowed us to include the share of correct answers as a control in regressions in which error types were the outcome of interest.[1] For example, the same overall score obtained by two participants could be obtained by solving different items, with different levels of difficulty and/or making different errors, so that, without matching, any difference observed in the proportion and type of errors made could be due to differences in overall performance and the characteristic qualities of the group (Facon & Nuchadee, 2010).

Within this framework, the objective of this study is to analyze the performance and error patterns of bilingual children aged 5 to 12, matched with their monolingual peers (Spanish and monolingual indigenous language speakers) with respect to the level of exposure to poverty,[2] certain individual characteristics (sex and age) and the type of area in which the household is located.[3] Specifically, our main research questions and hypotheses are the following:

- First, when considering the total available sample of children aged 5 to 12, after matching based on the level of exposure to poverty, sex, age, and type of household area do we observe differences in overall performance (number of correct answers) in the Raven's CPMT for the bilingual group with respect to their monolingual peers? Although the empirical evidence is not conclusive, according to the argument presented by Ben-Zeev (1977a) regarding a greater scanning ability by bilinguals, we hypothesized that bilingual children would have better performance (higher number of correct answers) on the CPMT than their monolingual peers in Spanish and their monolingual peers in an indigenous language (Hypothesis 1).
- Second, considering the total sample (children aged 5 to 12), once the participants were matched according to level of exposure to poverty, gender, age, and the type of household area, do bilingual and monolingual children make the same error patterns when offering an incorrect response to any item on the task? It was expected that bilingual children would make fewer Difference (D) errors than their monolingual peers due to their use of a different exploration

[1] We did not include the "Total Score of Raven's CPMT" in the model for the propensity score to avoid inducing a specific bias arising when including an outcome that could select individuals (as the error types could be affected by the independent variable of interest directly and indirectly through the total score).

[2] Taking advantage of the large amount of data provided by the MxFLS on the characteristics of households and the activities carried out by their members, in what we consider a modest contribution of this work, we have constructed a multidimensional poverty measure that includes factors associated with both material and symbolic poverty.

[3] The variable "type of household area" is a proxy to identify the type of area (locality) in which the child's household is located, based on the level of basic services it offers. It is a dichotomous variable that takes the value 1 if the household is located in a locality without basic sanitation infrastructure (for the evacuation of excreta), and 0 if there is any infrastructure (septic tank, sewage system, etc.).

strategy (Ben-Zeev, 1977b), but no differences were anticipated on the other error types (WP, R, IC) (Hypothesis 2).

– Finally, differentiating between two age cohorts for each of the linguistic groups in the sample (aged 5 to 8 and 9 to 12, respectively), after matching the participants based on level of exposure to poverty, sex, age, and the type of household area, is the pattern of errors maintained throughout primary education, regardless of the linguistic group? Here, according to the pattern described in typically developing children (Gunn & Jarrold, 2004; Petretto et al., 2021; Raven, Court, & Raven, 1990; Van Herwegen et al., 2011), we expected that the most common errors would be type R and the least frequent would be type D, for both bilinguals and monolinguals, and that this pattern of errors would be maintained in both age range groups (Hypothesis 3).

2. Method

2.1. Participants

The sample was obtained from participants in the third round of data collection (the last round available) of the Mexican National Survey on Household Living Standards (*Encuesta Nacional sobre Niveles de Vida de los Hogares - MxFLS*) (Rubalcava & Teruel, 2013). It consisted of 4644 primary school children aged 5 to 12 ($M = 8.56$, $SD = 2.25$, Min = 5, Max = 12), of which 49.71% were boys and 50.29% were girls.

Two age groups were established, the first containing 2256 (48.58%) children aged 5 to 8 ($M = 6.54$, $SD = 1.11$, Min = 5, Max = 8) and the second with 2388 (51.42%) children aged 9 to 12 ($M = 10.46$, $SD = 1.11$, Min = 9, Max = 12). The MxFLS is a multi-thematic and representative longitudinal survey of the Mexican population at the national, urban, rural, and regional levels. The MxFLS sampling design carried out by the National Institute of Statistics and Geography (INEGI, based on its Spanish name - *Instituto Nacional de Estadística y Geografía*) is probabilistic, stratified, multistage, and clustered. Participants who, according to an interviewer's report, had poor or very poor attention during the test were excluded from the original sample.

Three participant groups have been defined, based on the regular use of one or more languages: bilingual, monolingual in Spanish, and monolingual in an indigenous language. The distinction between bilingual and monolingual is established based on reporting by the head of the household in the MxFLS as to whether the child speaks Spanish and some indigenous language (bilingual), speaks only Spanish (monolingual in Spanish), or speaks only an indigenous language (monolingual in that language). Table 2 shows the distribution of participants in the three linguistic groups and the corresponding descriptive statistics according to sex, age and type of household area.

2.2. Instruments and measures

2.2.1. Cognitive performance

A reduced version of the Raven's CPMT (; Raven et al., 2006) was administered. The Raven's CPMT is a classic, non-verbal abstract reasoning test based on graphical analogy problem-solving, having good psychometric properties across diverse cultural, economic, and social settings.

Each problem consists of a matrix with figural elements, such as geometric figures and lines, in which an element is missing. Participants must look across the rows and then look down the columns to discover the rules governing the presentation of the diverse figural elements and then use these rules to determine the missing element. Participants must identify the relevant features of a set of visual elements and then select the correct element from six response alternatives arranged below the matrix. Although the test consists of three series of twelve items (series A, AB, and B), in the MxFLS, a reduced version containing 18 items of the Raven's CPMT was administered: 6 correspond to series A (A2, A4, A6, A8, A10, A12), 6 to the AB series (AB 2, AB 4, AB 6, AB 8, AB 10, AB 12) and 6 to series B (B2, B4, B6, B8, B10, and B12). This selection strategy is considered adequate since some studies have shown that completing the entire test can lead to a fatigue effect in young children (Van de Vijver & Brouwers, 2009). The following measures were obtained from the answers given to each item and from the test as a whole:

– *Total score on the Raven's CPM.* The total score on the Raven's CPMT was calculated as the number of correctly solved items.
– *Error types.* According to the classification of the distractors incorporated by each item of the test (Raven et al., 2006), four variables were defined, based on the types of conceptual error underlying the incorrect response to the item:
– *Repetition Errors (R).* The score was calculated as the number of incorrectly solved items in which the selected answer or element repeats the drawing that is above and to the left of the space to be completed in the matrix, or it is a figure that repeats the drawing that is above or to the left of the space to be filled.
– *Difference Errors (D).* The score was calculated as the number of incorrectly solved items in which the selected answer corresponds to an element that does not contain any figure or is an element that is completely incongruent with the given matrix.
– *Wrong Principle Errors (WP).* The score was calculated as the number of incorrectly solved items in which the chosen answer contains a figure that is contaminated by inconsistencies or distortions, combines figures incongruously, or is all or half of the drawing to be completed.
– *Incomplete Correlate Errors (IC).* The score was calculated as the number of incorrectly solved items in which the selected answer contains a figure that is misoriented or is an incomplete but correct figure.

2.2.2. Poverty

Based on a set of socioeconomic indicators included in the MxFLS, a *Synthetic index of global poverty* was calculated yielding values between 0 (more impoverished) and 10 (less impoverished). To calculate it, the scores obtained in two dimensions or sub-indices of poverty, material and symbolic, were added:

– *Material poverty* refers to restrictions on economic resources to meet a series of basic needs (Feres & Mancero, 2001). In this dimension, four dichotomous indicators (1 = yes; 0 = no) were included, corresponding to whether the child's habitual residence was endowed with certain characteristics (i.e., electricity, floor with insulating material, kitchen and bedroom in different rooms, WC utilities linked to the sanitation network) that could directly affect their physical well-being and health. This sub-index yields values between 0 (more material poverty) and 4 (less material poverty).
– *Symbolic poverty* refers to time restrictions and opportunities for the child to participate in certain activities mediated by verbal communication (oral and written) outside of the school environment (Sen, 2000), and that could contribute to his/her cognitive and linguistic enrichment. This sub-index yields values between 0 (highest level of symbolic poverty) and 6 (lowest level of symbolic poverty), based on the score obtained on six dichotomous indicators (1 = yes; 0 = no) revealing whether the child usually performs certain activities individually (i.e., reading, watching television and using the internet) or together with others (i.e., sports/culture, playing games inside or outside the home, helping other household members with study/homework).

2.3. Procedure

The reduced Raven's CPMT was administered individually in its pencil-and-paper version, presenting the 18 items one by one and in the same order of increasing difficulty with regard to the entire sample, as recommended by the manual (Raven et al., 2006). Each child was asked

Table 2
Group sample size and basic statistics according to sex, age and type of household area.

	Sample		Sex		Age				Type of household area	
Group	N	%	Boys	Girls	Mean	SD	Min	Max	Basic sanitation	No basic sanitation
Bilingual	671	14.20	335	336	8.66	2.22	5	12	643	28
Monolingual Spanish	3970	84.15	1970	2000	8.55	2.25	5	12	3826	132
Monolingual Indigenous	78	1.65	41	37	7.85	2.13	5	12	74	4
Total	4719	100	2346	2373	8.56	2.25	5	12	4543	164

Note: N = number of cases. % = percentage. Mean = arithmetic mean. SD = standard deviation. Min = minimum value. Max = maximum value. Source: Prepared by the authors using the data from the MxFLS database.

to choose the figure (from the 6 offered in each item) that completed the item matrix for each of the items presented. There was no time limit for completing the test and the child was encouraged to respond to all items. The test was administered in the child's home, with the interviewer recording the answers.

2.4. Data analyses

We conducted an assessment with a quasi-experimental multiple-treatment strategy. For the analysis of the results, Flores and Mitnik's methodological proposal (Flores & Mitnik, 2013) was followed. This consists of proposing a counterfactual which, for the purposes of our research, allows us to reduce the differences in the mean results of the comparison groups (bilingual vs. monolingual) due to individual differences in age, sex, exposure to material and symbolic poverty, and the type of household area. The counterfactual is initially based on the assumption of the subject selection hypothesis based on observable characteristics or "unconfoundedness" (Rubin, 1990); and secondly, it relies on non-experimental multiple treatment estimators to eliminate intergroup differences in observable characteristics.

Because propensity score matching only identifies the causal impact of an independent variable when the propensity score was created with all other causal variables or highly correlated with all other causal variables (Steiner, Cook, Shadish, & Clark, 2010), and because other variables may cause differences between groups (as previously discussed in the introduction), even after matching some of them via propensity scores, it is important to note the following: (1) any remaining differences between groups on the Raven's after matching represent the maximum possible causal effect; (2) the matching of certain variables may not eliminate their causal impact if they have indirect causes (e.g., interaction effects) on the dependent variables and/or share a common cause with the dependent variables; and (3) uninvestigated variables could cause the differences seen in this study.

This methodological proposal is implemented in three stages:

1) Estimation of the Generalized Propensity Score (GPS) (Imbens & Wooldridge, 2009; Hirano & Imbens, 2004). Through a multinomial logit model, the GPS estimates the probability that individual i belonging to group G is in location k, $\Pr(G_i = k | G_i = k)$. Likewise, for any individual that is not part of group G, it offers the probability that they belong to k, $\Pr(G_i = k | G_i \neq k)$ based on the observable characteristics considered. This last probability is relevant for the second stage, while the first probability is used in the third stage.
2) Establishment of an intergroup "common support condition", by virtue of which individuals are selected based on their observable characteristics, which overlap between groups so as to permit comparison. To do so, we select individuals whose probability of belonging to group k, given that they are not part of group k, is greater than the q^{th} quantile of the lowest probability of belonging to group k for k's individuals for each group. In this way, an individual satisfies the overlap condition in each group k, ensuring comparability based on observable characteristics. We followed Flores and Mitnik (2013) recommendation of selecting the quantile q that determines the amount of q = 0.0025. The use of a value of the q^{th} quantile that is greater than zero avoids the selection of an "outlier" that implies incorporating non-comparable individuals between the groups.[4]
3) Estimation of the measure associated with each group and analysis of intergroup differences. For this, a regression model $Y_{ik} = \sum_{k=1}^{K} \beta_k G_{ik} + \gamma X_i + \varepsilon_i$ is applied, where Y_{ik} is the number of responses of a given type;β_k are the coefficients of the regressors associated with each group k; X_i is the individual characteristics matrix; and ε_i is the error term. From this equation, three alternative estimators are calculated for the mean associated with each group:
   - *Linear Regression* (Raw Means). This is an estimator that calculates the expectation of the outcome conditional to its belonging to a certain group and certain individual characteristics $E[Y_{ik} | G_i = k]$. For each group, the mean of the estimated values is provided, without imposing any common support conditions.
   - *Linear Regression Imposing Overlap* (Raw Means overlap). This estimator provides the mean of the estimated values considering only the scores of individuals who satisfy the imposed overlap condition.
   - *Inverse Probability Weighting* (IPW). Considering the individuals that satisfy the imposed overlap condition, this estimator provides the mean of the inverse probability of its own GPS, $\Pr(G_i = k | G_i = k)$. Inverse probability weighting allows missing data to be accounted for when individuals with missing data cannot be included in the primary analysis (Hernán & Robins, 2006), thereby increasing the weight of underrepresented cases in the sample.

3. Results

Table 3 presents the results of the descriptive analysis for the total sample before performing the overlap in the Raven's test, expressed as a percentage (% times 1) of correct answers and the different types of error considered. Statistics are also included for the overall measurement of the poverty index, as well as its main dimensions (material and symbolic).

3.1. Overlap analysis

First, we present the results of the overlap between the different linguistic groups, for both the total sample and differentiated between two age groups: from 5 to 8 and from 9 to 12, with respect to their level of exposure to poverty, certain individual characteristics (sex and age), and type of household area. Table 4 shows the percentage of cases eliminated from each linguistic group in the total sample and in each age group (from 5 to 8 and from 9 to 12), after imposing the overlap, as well

[4] No consensus exists, even in the binary treatment literature, as to how to select the appropriate amount of trimming (Imbens & Wooldridge, 2009). Therefore, we coincide with Flores and Mitnik (2013), who established this general criterion to prevent a single individual from remaining inside or outside the overlap region. As a result, in the case of discontinuity in the tail of the overlap distribution (due to an outlier), that individual is excluded when determining the overlap region.

Table 3
Descriptive statistics results and poverty measures for the total sample.

	Mean	SD	Min	Max
Total Score RCPMT	0.57	0.21	0	1
Error types				
Difference	0.02	0.05	0	0.39
Repetition	0.28	0.15	0	0.67
Wrong Principle	0.07	0.08	0	0.50
Incomplete Correlate	0.10	0.08	0	0.50
Poverty				
Material	3.74	0.62	0	4
Symbolic	2.34	0.85	0	6
Global	6.08	1.01	0	10

Note: Mean = arithmetic mean. SD = standard deviation. Min = minimum value. Max = maximum value. Source: Prepared by the authors using the data from the MxFLS database.

Table 4
Group sample size with and without overlap.

N	Bilingual	Spanish	Indigenous	Total
Entire sample (children aged 5–12)				
n	549	3596	41	4186
n after imposing overlap	505	3312	41	3858
% cases dropped due to non-overlap	8.01%	7.90%	0.00%	7.84%
Children between 9 and 12				
n	289	1862	15	2166
n after imposing overlap	261	1659	15	1935
% cases dropped due to non-overlap	9.69%	10.90%	0.00%	10.66%
Children between 5 and 8				
n	260	1734	26	2020
% after imposing overlap	253	1688	26	1967
% cases dropped due to non-overlap	2.69%	2.65%	0.00%	2.62%

Note: N = number of cases. % = percentage. Source: Prepared by the authors using the data from the MxFLS database.

as the total percentage of eliminated cases (7.84%). It is evident that in the group of monolingual Spanish-speaking children, aged 9 to 12 years, almost 11% of the cases were discarded since they were not comparable with children of at least one of the other linguistic groups (bilingual and monolingual in an indigenous language).

Fig. 1 graphically shows the quality of the overlap carried out (based on the *GPS* distribution), for both the total sample (panel a), and for the subsamples of children aged 9–12 (panel b) and 5–8 (panel c).

3.2. GPS estimation and covariate balancing

The GPS balances the covariates between individuals belonging to a certain group and those who do not belong to it (Imbens, 2000). To examine the balance of each covariate after adjusting the GPS, Flores and Mitnik's strategy (Flores & Mitnik, 2013) was applied to determine whether there is equality of joint means after each observation is weighted by $1/\widehat{R}_i$. To estimate the GPS, a *multinomial logit (MNL)* model including 4 covariates (synthetic global poverty index, age, sex, and type of household area) was used. Table 5 presents for the entire sample, the socioeconomic (poverty level), individual (sex and age) and household area characteristics (type of household area), both without applying "overlap" (without equalization) and while performing an "overlap" analysis (equated or with common support). In addition, the differentiated results are presented in two age cohorts, from 5 to 8 and from 9 to 12. Estimation results are presented for each of the participating groups (bilingual, monolingual in Spanish, and monolingual in an indigenous language) for the total sample and for each of the two age cohorts. Specifically, the *p*-value of the Root-Mean-Square Deviation (*RMSD*) test (see equation - 3 - of the methodological appendix) is shown for the mean of each variable calculated before and after weighting by the GPS.

The results show how the weighting by GPS brings the means of all the covariates closer for the three linguistic groups, both when considering the whole sample and the subsamples of the two age ranges, with the only exceptions of the type of household area (in the total sample and in the group of children aged 5 to 8), and sex (in the group of children aged 9–12). For example, the RMSD estimator shows that the global poverty index in the entire sample (5–12 years of age) is reduced from 0.025 to 0.013 after weighting the GPS. This significant improvement in the balance of the three language groups is evident when considering the joint test of equality of means, which is only rejected for the sex covariate. Thus, we find that GPS weighting greatly improves the balance between the three language groups. We tested for mean differences of the variables used in our regressions. However, age and the constructed poverty index are unbalanced in the full sample.[5] After imposing the common support and inverse probability weighting matching, we did not find statistically significant differences between the means of the three groups (see, Table 5, *p*-values). Therefore, based on these results, we can conclude that, the covariates in the raw data were based on a considerable balance between the three language groups, and the estimated GPS does a good job of improving their balance in the entire sample (5–12 years of age).

3.3. Main results

3.3.1. Estimation without matching (overlap)

Table 6 presents the estimated mean values without matching (overlap) for the overall performance on the Raven's CPMT (Total score RCPMT) and the percentage of each type of error variable for each language group and age cohort. When considering the total sample, bilingual children are found to have a significantly lower overall performance on the Raven's CPMT as compared to monolingual (Spanish-speaking) children, without making the comparison based on the variables of sex, age, poverty level and type of household area. When considering different age cohorts, we see that both bilingual children aged 9 to 12 and younger (aged 5–8) children revealed a tendency to underperform their monolingual (Spanish-speaking) peers.

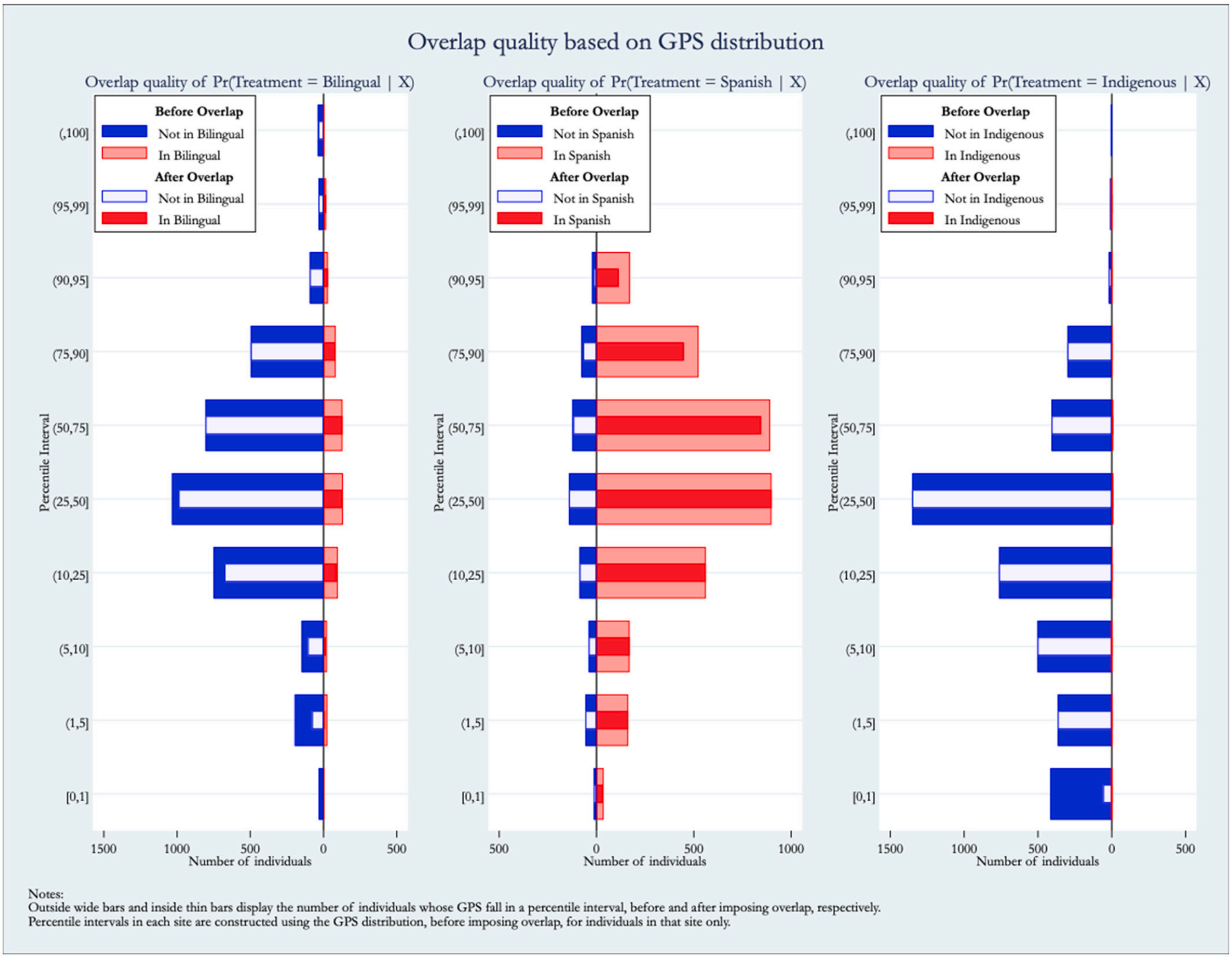
When comparing bilinguals and monolinguals, no statistically significant differences were found in the frequency of any of the error types when solving the Raven's CPMT, indicating that the intergroup differences are quantitative and not qualitative. Finally, when individually analyzing the error patterns of each linguistic group, we find that, regardless of whether they are bilingual or monolingual (either in Spanish or an indigenous language), the most frequent type of error made by children aged between 5 and 8 was R, followed by IC and WP, whereas the least frequent error was type D. Likewise, in older children (9–12 years), this same pattern of relative frequency of the four error types was maintained.

Additionally, considering the two age cohorts, Table 7 shows how the intragroup differences in each linguistic group, are only statistically significant in the case of errors type D, in the group of monolingual children in Spanish.

3.3.2. Estimation with matching (overlap)

Table 8 shows the mean values estimated using a linear regression model where the condition of common support (overlap) was imposed with respect to the variables of age, sex, poverty level and type of household area for each language and age group, in order to analyze the

[5] In these cases, the propensity score matching did not fully control for the differing poverty among the three groups. In other words, at least some of the remaining group differences could be due to remaining effects of poverty and age.



(a) All children (5-12 years of age)

Fig. 1. Overlap Quality-Plots (based on GPS distribution).
(a) All children (5–12 years of age)
(b) Older children (9–12 years of age)
(c) Younger children (5–8 years of age)
Source: Prepared by the authors using the data from the MxFLS database.

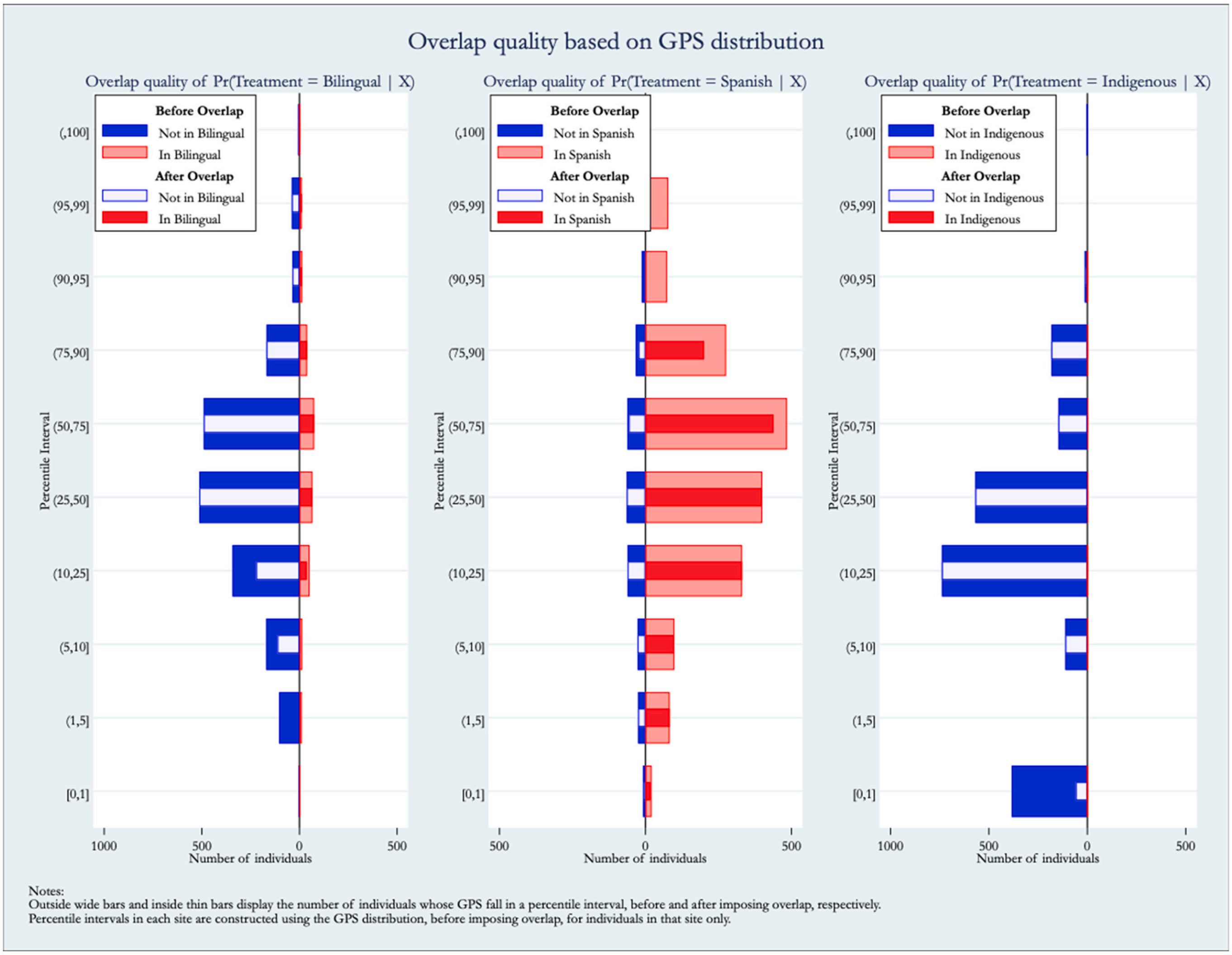
intergroup differences of the overall performance on the Raven's test (Total score RCPMT). When considering the total sample, the results are found to be consistent in both estimates, with and without overlap, although the mean values of the overall performance on the Raven's CPMT were slightly lower when the estimate was carried out with overlap. Again, the results reveal that bilingual children had a significantly lower overall performance on the CPMT than monolingual (Spanish-speaking) children. Moreover, when considering different age cohorts, we observe that both the older (9–12-year-olds) and younger (5–8-year-olds) bilinguals had a lower performance than monolingual Spanish speakers, although this difference was not statistically significant.

Regarding the pattern of errors on the Raven's CPMT, the results indicate that there were no statistically significant differences between groups in the frequency with which they tended to make each type of error. The results obtained when imposing overlap are consistent with the previous estimation in which common support was not imposed. Thus, regardless of the bilingual or monolingual condition in the younger children (5–8 years), the most frequent type of error was R, followed by IC, WP, and, finally, D, maintaining this pattern in older children. Type D errors had the same frequency for children aged 5 to 12.

Again, considering the two age cohorts, the Table 9 results are consistent with the previous estimator to reveal how the intragroup differences in each linguistic group, are only statistically significant in the case of errors type D, in the group of monolingual children in Spanish.

3.3.3. *Estimation with overlap and inverse probability weighting*

Table 10 presents a third estimator for each group, inverse probability weighting (IPW), which permits the increased weighting of monolingual children (in some indigenous language) who were underrepresented in the initial sample. When considering the total sample, once the overlap had been imposed, the results were consistent with those obtained in the previous estimation, although the mean values with respect to the overall performance on the Raven's CPMT were



(b) Older children (9-12 years of age)

Fig. 1. (*continued*).

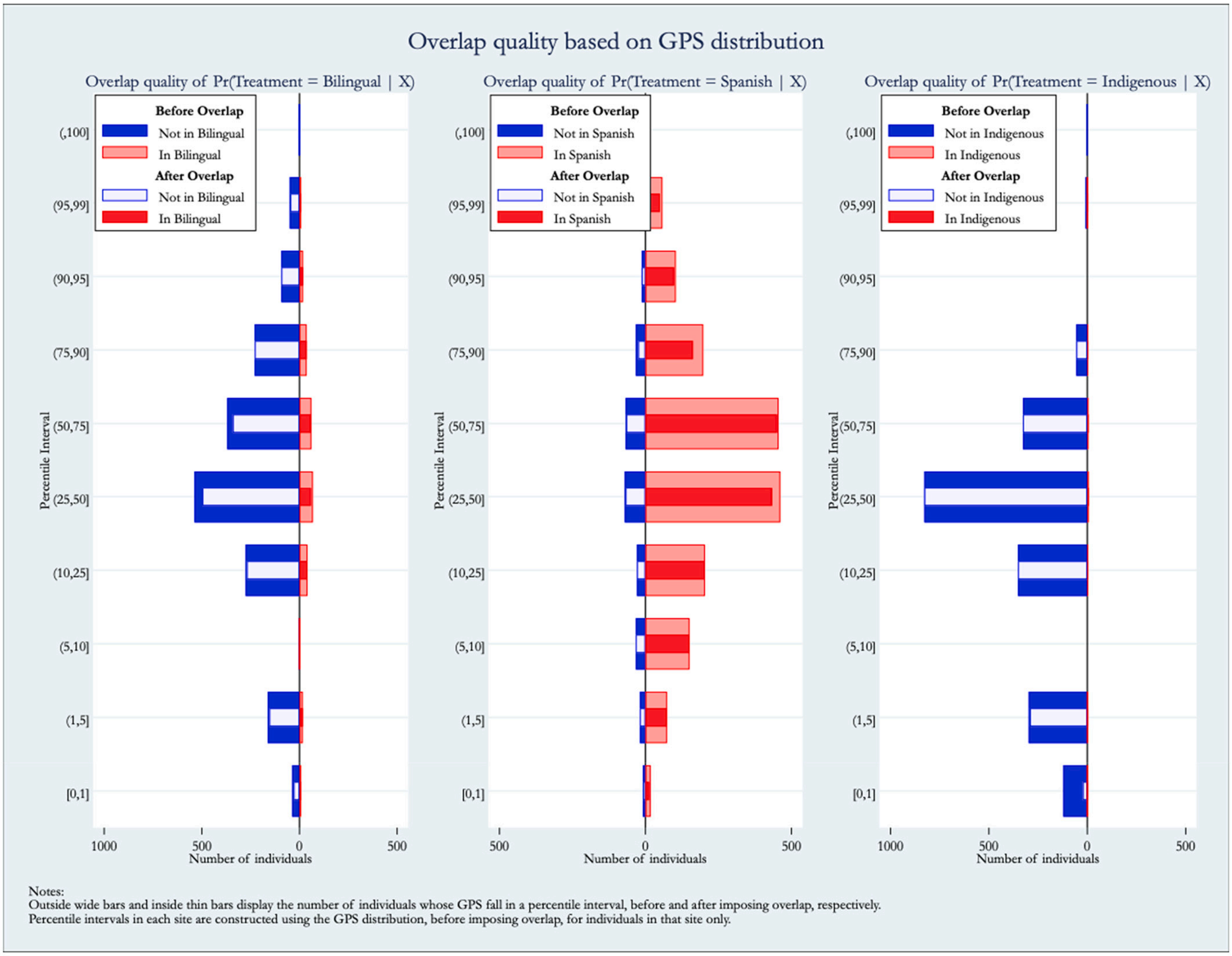
slightly lower in each of the groups. Again, bilingual children had a significantly lower overall performance on the test than monolingual (Spanish-speaking) children. Furthermore, when considering different age cohorts, we find that both older bilingual children (9–12-year-olds) and younger bilingual children (5–8-year-olds) performed significantly lower than the monolingual (Spanish-speaking) children.

Finally, it may be observed that when considering the different types of error, as in the two previous estimates, no statistically significant differences were obtained between linguistic groups. Likewise, the results obtained from each of the groups were consistent with previous estimates, with and without common support overlap, whereby children aged 5–8 tended to make R errors more frequently, followed by IC, and WP, with D errors being the least frequently made. This same pattern of relative frequencies in error types was observed for children up to the age of 12.

Lastly, considering the two age cohorts, the Table 11 results are consistent with the two previous estimators to reveal how the intragroup differences in each linguistic group, are only statistically significant in the case of errors type D, in the group of monolingual children in Spanish.

4. Discussion

Using a representative sample of the Mexican population, the main objective of this study was to analyze performance on the Raven's CPMT by children aged 5 to 12 living in a multi-ethnic environment of poverty. Quantitative (correct answers) and qualitative (types of error) terms were used to compare the performance of three participant groups (monolinguals in Spanish, monolinguals in an indigenous language, bilinguals in Spanish and indigenous languages) on the Raven's task. Moreover, as a contribution of methodological approaches based on the comparison of groups, GPS techniques were used, allowing for the equating of participants in terms of variables such as age, sex, type of household area and level of exposure poverty (material and symbolic). In this sense, new evidence is provided on the relationships between the use of one or more languages and cognitive performance on a widely used non-verbal reasoning test. First, the results showed that significant differences exist between bilinguals and monolinguals in overall performance on the Raven's CPMT, with monolingual (Spanish-speaking) children achieving significantly higher scores than bilinguals. Secondly, we verified that, throughout primary education ages, no differences are found in the pattern of errors made by bilingual and monolingual (Spanish or indigenous language speakers) when completing the test. The main contributions in this regard, their limitations and necessary



(c) Younger children (5-8 years of age)

Fig. 1. (*continued*).

caution when interpreting the results, and their implications are presented below.

The results of the overall performance on the Raven's CPMT, show that, even with the comparison of groups being carried out with GPS techniques, significant differences exist between bilingual and monolingual (Spanish-speaking) children, with the latter earning higher scores from the ages of 5 to 12, a period coinciding with primary school enrollment. Therefore, Hypothesis 1 was not supported. We provide a more detailed analysis whose results complement those reported in previous studies in which the superior performance of bilingual children as compared to monolingual ones was either only found under conditions of poverty (Ben-Zeev, 1977a) or was not supported (Bialystok & Shorbagi, 2021; Engel de Abreu et al., 2012). Our findings allow us to qualify prior evidence that revealed a significantly lower performance by indigenous children on the RCPMT as compared to their non-indigenous counterparts (Millones et al., 2015; Nistal, 2014). In this study, once participants have been equated in terms of age, sex, type of household area and level of exposure to poverty (material and symbolic), Spanish monolinguals (non-indigenous) are found to be the most advantaged children on overall task performance, whereas some factors such as belonging to a minority ethnic community appear to negatively influence performance on non-verbal reasoning tests such as the Raven's CPMT, particularly in the case of bilinguals.

Some studies have found that bilingual children from minorities and low-income families did not differ from their monolingual peers in terms of cognitive abilities such as abstract reasoning (Engel de Abreu et al., 2012) and attentional control (Kalashnikova, Pejovic, & Carreiras, 2021). It was also reported that the heritability of intelligence, specifically Raven's scores as indicators of IQ (intelligence quotient), does not differ across ethnic groups living in disadvantaged environments (Pesta, Kirkegaard, Nijenhuis, Lasker, & Fuerst, 2020). However, it has not been ruled out that monolingual intergroup differences could emerge based on linguistic experiences at home (Kroll & Bialystok, 2013a, 2013b) and at school (Woumans, Surmont, Struys, & Duyck, 2016), parental education (Millones, Flores-Mendoza and Rivalles, 2015), and degree of family stimulation practices at home (Knauer, Ozer, Dow, & Fernald, 2018). In other words, the environmental quality could play a meaningful role in explaining differences in Raven's performance between the groups of children (Lynn et al., 2018). Spanish was the language used by interviewers when administering the Raven's CPMT test to participants in this study. It is also the majority social language in Mexico and the main language of literacy instruction for both indigenous and non-indigenous children. Hence, the higher performance by monolinguals (Spanish speakers) as compared to that of bilinguals may be partly attributed to the fact that indigenous bilinguals face significant challenges in terms of language barriers.

Table 5
Descriptive statistics for overlap of the groups and evaluation of the estimators.

														Joint Means					Joint Equality of Mean Wald Test			Joint Root Mean Square Distance		
	Raw Mean and SD without overlap								Raw Mean and SD with overlap					Raw Mean and SD IPW										
	Bilingual		Spanish		Indigenous		Bilingual		Spanish		Indigenous		Bilingual		Spanish		Indigenous		*p*-value			RMSD		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	Raw	Raw v/ov	IPM	Raw	Raw v/ov	IPM
Entire sample (5–12 years of age)																								
Age	8.64	2.21	8.56	2.26	7.78	2.13	8.42	2.15	8.35	2.18	7.78	2.13	8.33	2.15	8.35	2.19	8.57	2.18	0.042	0.173	0.815	0.047	0.035	0.013
Sex	0.48	0.50	0.50	0.50	0.49	0.51	0.48	0.50	0.49	0.50	0.49	0.51	0.50	0.50	0.49	0.50	0.57	0.50	0.758	0.885	0.665	0.014	0.010	0.068
Poverty	6.05	1.07	6.11	0.98	5.77	1.18	5.88	0.91	5.96	0.81	5.77	1.18	5.92	0.90	5.95	0.81	6.07	0.92	0.098	0.126	0.507	0.025	0.013	0.011
Area	0.96	0.19	0.97	0.17	0.90	0.30	0.96	0.20	0.97	0.17	0.90	0.30	0.97	0.18	0.97	0.18	0.97	0.17	0.219	0.199	0.907	0.032	0.031	0.003
Children between 9 and 12																								
Age	10.47	1.08	10.46	1.11	10.13	1.13	10.33	1.04	10.30	1.06	10.13	1.13	10.30	1.04	10.31	1.06	10.39	1.14	0.502	0.754	0.772	0.015	0.009	0.010
Sex	0.48	0.50	0.49	0.50	0.47	0.52	0.47	0.50	0.48	0.50	0.47	0.52	0.48	0.50	0.48	0.50	0.56	0.51	0.966	0.948	0.715	0.018	0.011	0.099
Poverty	6.15	1.17	6.26	1.06	6.00	0.99	5.97	1.06	6.07	0.92	6.00	0.99	6.03	1.06	6.06	0.92	6.17	0.80	0.227	0.352	0.791	0.017	0.007	0.009
Area	0.96	0.19	0.97	0.17	0.93	0.26	0.96	0.20	0.97	0.18	0.93	0.26	0.96	0.19	0.97	0.18	0.98	0.16	0.662	0.673	0.938	0.017	0.015	0.005
Children between 5 and 8																								
Age	6.61	1.09	6.52	1.11	6.42	1.14	6.59	1.10	6.50	1.11	6.42	1.14	6.57	1.10	6.51	1.11	6.64	1.07	0.450	0.450	0.962	0.012	0.011	0.004
Sex	0.48	0.50	0.51	0.50	0.50	0.51	0.49	0.50	0.51	0.50	0.50	0.51	0.50	0.50	0.51	0.50	0.57	0.50	0.703	0.796	0.835	0.023	0.019	0.051
Poverty	5.93	0.93	5.94	0.85	5.63	1.27	5.87	0.84	5.89	0.76	5.63	1.27	5.90	0.82	5.89	0.76	5.97	1.04	0.444	0.503	0.766	0.025	0.020	0.010
Area	0.96	0.19	0.97	0.17	0.88	0.33	0.96	0.20	0.97	0.17	0.88	0.33	0.97	0.18	0.97	0.17	0.97	0.18	0.323	0.286	0.975	0.041	0.041	0.002

Note: M = arithmetic mean. SE = standard error. ICi = lower confidence interval. ICs = upper confidence interval. B = Bilingual. S = monolingual in Spanish. I = monolingual in an indigenous language. Poverty = Synthetic index of global poverty; Area = type of household area. The table shows the *p*-values of the joint equality test in Eq. (2) and the RMSD in Eq. (3) of the methodological appendix, with the corresponding 95% confidence interval based on a "bootstrap" of 1000 replicates and the estimates and confidence intervals of βd for each group. Source: Prepared by the authors.

Table 6
Wald test: first estimator Reg. Adj. Mean.

	Bilingual (B)				Spanish (S)				Indigenous (I)									Intergroup Comparisons *p*		
Variable	M	SE	ICi	ICs	M	SE	ICi	ICs	M	SE	ICi	ICs	*p* value	Effect size	RMSD	ICi	ICs	B-S	S-I	B-I
Entire sample (5–12 years of age)																				
Total Score RCPMT	0.54	0.0082	0.52	0.56	0.58	0.0035	0.57	0.59	0.54	0.0267	0.49	0.59	0.980	0.254	0.033	0.01	0.06	< 0.001	0.143	0.984
Error types																				
Difference	0.02	0.0020	0.02	0.02	0.02	0.0007	0.02	0.02	0.02	0.0076	0.01	0.03	0.970	0.040	0.040	−0.17	0.25	0.479	0.819	0.967
Repetition	0.27	0.0046	0.26	0.28	0.28	0.0022	0.27	0.28	0.27	0.0159	0.24	0.30	0.970	0.149	0.008	−0.02	0.04	0.357	0.762	0.972
Wrong Principle	0.07	0.0028	0.07	0.08	0.07	0.0013	0.06	0.07	0.07	0.0096	0.05	0.09	0.730	0.131	0.026	−0.05	0.10	0.180	0.963	0.732
Incomplete Correlate	0.10	0.0027	0.09	0.10	0.10	0.0013	0.09	0.10	0.10	0.0100	0.08	0.12	0.460	0.095	0.036	−0.03	0.10	0.869	0.469	0.459
Children between 9 and 12																				
Total Score RCPMT	0.64	0.0113	0.61	0.66	0.67	0.0041	0.66	0.68	0.58	0.0480	0.48	0.67	0.210	0.076	0.064	0.01	0.13	0.002	0.046	0.207
Error types																				
Difference	0.01	0.0020	0.01	0.02	0.01	0.0008	0.01	0.01	0.02	0.0110	0.01	0.04	0.480	0.009	0.264	−0.18	0.71	0.843	0.448	0.479
Repetition	0.22	0.0059	0.21	0.23	0.22	0.0032	0.22	0.23	0.21	0.0257	0.16	0.26	0.710	0.070	0.029	−0.05	0.11	0.342	0.558	0.713
Wrong Principle	0.05	0.0032	0.04	0.05	0.04	0.0014	0.04	0.05	0.05	0.0106	0.03	0.07	1.000	0.020	0.034	−0.09	0.16	0.342	0.759	0.996
Incomplete Correlate	0.08	0.0037	0.07	0.08	0.07	0.0017	0.07	0.08	0.09	0.0159	0.06	0.12	0.470	0.016	0.078	−0.04	0.20	0.488	0.367	0.475
Children between 5 and 8																				
Total Score RCPMT	0.44	0.0132	0.41	0.46	0.48	0.0046	0.47	0.49	0.48	0.0326	0.41	0.54	0.280	0.059	0.041	0.01	0.08	0.002	0.871	0.279
Error types																				
Difference	0.03	0.0035	0.02	0.03	0.02	0.0012	0.02	0.03	0.02	0.0103	0.01	0.04	0.650	0.042	0.085	−0.18	0.34	0.399	0.842	0.650
Repetition	0.33	0.0067	0.31	0.34	0.33	0.0034	0.32	0.34	0.33	0.0223	0.29	0.37	0.900	0.005	0.006	−0.03	0.04	0.544	0.960	0.896
Wrong Principle	0.10	0.0047	0.09	0.11	0.09	0.0021	0.09	0.10	0.09	0.0139	0.07	0.12	0.770	0.068	0.025	−0.05	0.10	0.264	0.936	0.766
Incomplete Correlate	0.12	0.0040	0.11	0.12	0.12	0.0020	0.12	0.12	0.12	0.0143	0.09	0.15	0.620	0.031	0.026	−0.04	0.09	0.373	0.792	0.618

Note: M = arithmetic mean. SE = standard error. ICi = lower confidence interval. ICs = upper confidence interval. B = Bilingual. S = monolingual in Spanish. I = monolingual in an indigenous language. Total Score RCPMT = number of correct answers on the Raven's test; Difference = frequency of difference errors; Incomplete Correlate = frequency of errors due to incomplete correlate; Repetition = frequency of repetition errors; Wrong Principle = frequency of errors due to incorrect individualization. The table shows the p-values of the joint equality test in Eq. (2) and the RMSD in Eq. (3) of the methodological appendix, with the corresponding 95% confidence interval based on a "bootstrap" of 1000 replicates and the estimates and confidence intervals of βd for each group. Source: Prepared by the authors.

Table 7
Wald test: first estimator Reg. Adj. Mean.

	Children between 5 and 8				*Children between 9 and 12*								
Variable	M	SE	ICi	ICs	M	SE	ICi	ICs	*p* value	Effect size	RMSD	ICi	ICs
All													
Total Score RCPMT	0.48	0.0042	0.47	0.49	0.66	0.0041	0.66	0.67	< 0.001	0.197	0.158	0.15	0.17
Error types													
Difference	0.01	0.0009	0.01	0.02	0.02	0.0011	0.02	0.02	< 0.001	0.025	0.238	0.16	0.32
Repetition	0.28	0.0027	0.27	0.28	0.27	0.0029	0.26	0.28	0.029	0.126	0.014	0.01	0.03
Wrong Principle	0.07	0.0015	0.07	0.07	0.07	0.0015	0.06	0.07	0.217	0.094	0.019	−0.01	0.04
Incomplete Correlate	0.10	0.0014	0.09	0.10	0.10	0.0017	0.09	0.10	0.962	0.075	0.001	−0.01	0.01
Bilingual													
Total Score RCPMT	0.45	0.0121	0.42	0.47	0.63	0.0121	0.61	0.66	< 0.001	0.198	0.172	0.14	0.20
Error types													
Difference	0.02	0.0029	0.01	0.02	0.03	0.0031	0.02	0.03	0.055	0.051	0.174	0.01	0.34
Repetition	0.29	0.0070	0.28	0.31	0.28	0.0082	0.27	0.30	0.350	0.124	0.016	−0.01	0.04
Wrong Principle	0.08	0.0050	0.07	0.09	0.08	0.0046	0.07	0.09	0.638	0.095	0.019	−0.03	0.07
Incomplete Correlate	0.10	0.0043	0.09	0.11	0.10	0.0051	0.09	0.11	0.562	0.069	0.018	−0.03	0.06
Spanish													
Total Score RCPMT	0.49	0.0045	0.48	0.50	0.67	0.0041	0.66	0.68	< 0.001	0.161	0.157	0.15	0.17
Error types													
Difference	0.01	0.0009	0.01	0.01	0.02	0.0012	0.02	0.02	< 0.001	0.067	0.242	0.16	0.33
Repetition	0.28	0.0029	0.27	0.28	0.27	0.0030	0.26	0.27	0.066	0.043	0.013	0.01	0.03
Wrong Principle	0.07	0.0017	0.06	0.07	0.06	0.0016	0.06	0.07	0.265	0.114	0.019	−0.01	0.05
Incomplete Correlate	0.09	0.0017	0.09	0.10	0.09	0.0018	0.09	0.10	0.785	0.085	0.004	−0.01	0.02
Indigenous													
Total Score RCPMT	0.45	0.0340	0.39	0.52	0.54	0.0598	0.42	0.65	0.227	0.198	0.085	−0.03	0.20
Error types													
Difference	0.02	0.0099	0.01	0.04	0.04	0.0144	0.01	0.07	0.187	0.023	0.331	−0.07	0.73
Repetition	0.32	0.0257	0.26	0.37	0.31	0.0297	0.25	0.36	0.807	0.128	0.014	−0.05	0.08
Wrong Principle	0.09	0.0154	0.06	0.12	0.09	0.0150	0.06	0.12	0.764	0.095	0.030	−0.10	0.16
Incomplete Correlate	0.12	0.0133	0.10	0.15	0.13	0.0211	0.09	0.17	0.838	0.077	0.020	−0.10	0.14

Note: M = arithmetic mean. SE = standard error. ICi = lower confidence interval. ICs = upper confidence interval. B = Bilingual. S = monolingual in Spanish. I = monolingual in an indigenous language. Total Score RCPMT = number of correct answers on the Raven's test; Difference = frequency of difference errors; Incomplete Correlate = frequency of errors due to incomplete correlate; Repetition = frequency of repetition errors; Wrong Principle = frequency of errors due to incorrect individualization. The table shows the p-values of the joint equality test in Eq. (2) and the RMSD in Eq. (3) of the methodological appendix, with the corresponding 95% confidence interval based on a "bootstrap" of 1000 replicates and the estimates and confidence intervals of βd for each group. Source: Prepared by the authors.

The fact that the results do not allow us to confirm our first hypothesis of higher performance by bilingual children as compared to their monolingual peers suggests that bilingualism did not provide an advantage on Raven's CPMT performance. This is not surprising if we consider the current debate or lack of consensus regarding the alleged advantage of bilingualism for certain cognitive functions (Antoniou, 2019). Regarding this, several authors have mentioned the possibility of confirmation biases and the accumulation of common research practices (Paap, Johnson, & Sawi, 2015), as well as publication biases (De Bruin, Treccani, & Della Sala, 2015), which make the possible real advantages of bilingualism questionable in terms of effect size and the circumstances in which its benefits take place. A meta-analysis performed by Lehtonen et al. (2018) concluded that, once the mentioned biases were corrected, the available evidence did not provide systematic support for the general idea that bilingualism is associated with benefits in cognitive control functions in adults. Even in cases in which bilingualism has been associated with greater ability in cognitive control functions, its potential impact on the Raven's task performance is not always clear. Thus, when comparing early bilingual and monolingual adults, Treccani, Argyri, Sorace, and Della Sala (2009) found that although balanced bilinguals show a greater ability to inhibit distracting or irrelevant information (e.g., distracting stimuli in the Raven's matrices), this ability may be advantageous in some conditions but disadvantageous in others, such as under negative priming conditions in which previously irrelevant information becomes relevant at another time. These findings raise the question as to the possible specific non-linguistic cognitive effects of bilingualism on inhibitory control functions, which are effects that would not necessarily translate into cognitive advantages for the resolution of non-verbal reasoning tasks.

Our approach and analyses have not specifically focused on the study of executive functions, which are associated with children's performance on fluid intelligence tests (Arán-Filippetti, Krumm, & Raimondi, 2015; Johann et al., 2022). However, since these functions could differentially affect the processing of information underlying the different strategies used to solve this type of reasoning task, we hypothesized that they could translate into differences in the error patterns observed when solving the Raven's CPMT. Our results showed that whether the comparison of groups is considered or if the proposed one is not carried out, no significant differences exist between groups in terms of the frequency and types of error made by children, both when considering the entire sample or each individual age cohort (5–8 and 9–12 years). Consequently, Hypothesis 2 was not supported. This, however, does not discard potential differential cognitive effects that could influence information processing. These effects would not necessarily be reflected in significant differences in the strategies used to resolve the items and the resulting pattern of erroneous responses on the Raven's task.

The comparison of our results with those from previous studies is complex given that significant differences exist in the level of equality of the groups based on their individual characteristics, level of poverty, and even their linguistic condition. For example, of the available studies in which differences have been observed in some of the error types, children with "typical" development have been considered in some cases, but without a specific comparison according to the conditions of poverty and without differentiating their language situation (Gunn & Jarrold, 2004; Matzen et al., 1994; Petretto et al., 2021; Van Herwegen et al., 2011). In other cases, bilingual and monolingual children were explicitly compared (Ben-Zeev, 1977a) even considering their level of poverty (Ben-Zeev, 1977b), but an adequate comparison was not carried out with respect to overall performance on the Raven's test, in addition to having relatively small sample sizes. In studies in which, as in our case, no significant differences are found in the error patterns of

Table 8
Wald test: second estimator Reg. Adj. Mean w/overlap.

	Bilingual (B)				Spanish (S)				Indigenous (I)									Intergroup Comparisons *p*		
Variable	M	SE	ICi	ICs	M	SE	ICi	ICs	M	SE	ICi	ICs	*p* value	Effect size	RMSD	ICi	ICs	B-S	S-I	B-I
Entire sample (5–12 years of age)																				
Total Score RCPMT	0.53	0.0113	0.50	0.55	0.57	0.0081	0.55	0.59	0.53	0.0293	0.47	0.59	0.900	0.236	0.037	0.01	0.06	< 0.001	0.142	0.900
Error types																				
Difference	0.02	0.0023	0.02	0.02	0.02	0.0009	0.02	0.02	0.02	0.0076	0.01	0.03	0.960	0.041	0.036	−0.16	0.23	0.582	0.829	0.957
Repetition	0.28	0.0067	0.26	0.29	0.28	0.0050	0.27	0.29	0.28	0.0168	0.24	0.31	0.990	0.131	0.008	−0.02	0.04	0.371	0.772	0.990
Wrong Principle	0.07	0.0035	0.07	0.08	0.07	0.0022	0.07	0.07	0.07	0.0095	0.05	0.09	0.760	0.125	0.021	−0.05	0.10	0.323	0.992	0.762
Incomplete Correlate	0.10	0.0032	0.09	0.10	0.10	0.0022	0.09	0.10	0.11	0.0101	0.09	0.13	0.490	0.087	0.034	−0.03	0.10	0.989	0.472	0.492
Children between 9 and 12																				
Total Score RCPMT	0.62	0.0158	0.59	0.66	0.67	0.0087	0.65	0.68	0.57	0.0520	0.46	0.67	0.230	0.070	0.066	0.01	0.13	0.003	0.044	0.233
Error types																				
Difference	0.01	0.0025	0.01	0.02	0.01	0.0011	0.01	0.01	0.02	0.0112	0.01	0.04	0.490	0.008	0.250	−0.20	0.70	0.950	0.472	0.488
Repetition	0.22	0.0097	0.20	0.24	0.23	0.0066	0.22	0.24	0.21	0.0247	0.17	0.26	0.770	0.064	0.027	−0.05	0.10	0.325	0.579	0.774
Wrong Principle	0.05	0.0037	0.04	0.05	0.04	0.0023	0.04	0.05	0.05	0.0108	0.03	0.07	0.990	0.018	0.029	−0.09	0.15	0.495	0.783	0.987
Incomplete Correlate	0.08	0.0044	0.07	0.09	0.08	0.0024	0.07	0.08	0.09	0.0164	0.06	0.12	0.540	0.014	0.073	−0.05	0.19	0.485	0.389	0.539
Children between 5 and 8																				
Total Score RCPMT	0.44	0.0151	0.41	0.47	0.48	0.0066	0.47	0.49	0.48	0.0345	0.41	0.54	0.320	0.057	0.039	0.01	0.08	0.008	0.904	0.318
Error types																				
Difference	0.03	0.0040	0.02	0.04	0.02	0.0015	0.02	0.03	0.02	0.0106	0.01	0.04	0.600	0.042	0.100	−0.14	0.34	0.338	0.832	0.603
Repetition	0.33	0.0074	0.31	0.34	0.33	0.0038	0.32	0.34	0.33	0.0233	0.28	0.37	0.870	0.005	0.006	−0.03	0.05	0.526	0.972	0.873
Wrong Principle	0.10	0.0053	0.09	0.11	0.09	0.0030	0.09	0.10	0.09	0.0145	0.06	0.12	0.780	0.065	0.023	−0.06	0.11	0.333	0.961	0.775
Incomplete Correlate	0.12	0.0044	0.11	0.12	0.12	0.0025	0.11	0.12	0.12	0.0147	0.09	0.15	0.600	0.031	0.028	−0.04	0.10	0.323	0.800	0.604

Note: M = arithmetic mean. SE = standard error. ICi = lower confidence interval. ICs = upper confidence interval. B = Bilingual. S = monolingual in Spanish. I = monolingual in an indigenous language. Total Score RCPMT = number of correct answers on the Raven's test; Difference = frequency of difference errors; Incomplete Correlate = frequency of errors due to incomplete correlate; Repetition = frequency of repetition errors; Wrong Principle = frequency of errors due to incorrect individualization. The table shows the p-values of the joint equality test in Eq. (2) and the RMSD in Eq. (3) of the methodological appendix, with the corresponding 95% confidence interval based on a "bootstrap" of 1000 replicates and the estimates and confidence intervals of βd for each group. Source: Prepared by the authors.

Table 9
Wald test: second estimator Reg. Adj. Mean w/overlap.

	Children between 5 and 8				*Children between 9 and 12*								
Variable	M	SE	ICi	ICs	M	SE	ICi	ICs	*p* value	Effect size	RMSD	ICi	ICs
All													
Total Score RCPMT	0.48	0.0042	0.47	0.49	0.66	0.0041	0.66	0.67	< 0.001	0.197	0.158	0.15	0.17
Error types													
Difference	0.01	0.0009	0.01	0.02	0.02	0.0011	0.02	0.02	< 0.001	0.025	0.238	0.16	0.32
Repetition	0.28	0.0027	0.27	0.28	0.27	0.0029	0.26	0.28	0.030	0.126	0.013	0.01	0.03
Wrong Principle	0.07	0.0015	0.07	0.07	0.07	0.0015	0.06	0.07	0.219	0.094	0.019	−0.01	0.04
Incomplete Correlate	0.10	0.0014	0.09	0.10	0.10	0.0017	0.09	0.10	0.952	0.075	0.001	−0.01	0.01
Bilingual													
Total Score RCPMT	0.45	0.0121	0.42	0.47	0.63	0.0123	0.61	0.66	< 0.001	0.190	0.172	0.14	0.20
Error types													
Difference	0.02	0.0029	0.01	0.02	0.03	0.0032	0.02	0.03	0.060	0.051	0.169	0.01	0.34
Repetition	0.29	0.0071	0.28	0.31	0.28	0.0083	0.27	0.30	0.377	0.117	0.015	−0.01	0.04
Wrong Principle	0.08	0.0050	0.07	0.09	0.08	0.0046	0.07	0.09	0.615	0.094	0.021	−0.03	0.07
Incomplete Correlate	0.10	0.0043	0.09	0.11	0.10	0.0051	0.09	0.11	0.589	0.067	0.017	−0.03	0.06
Spanish													
Total Score RCPMT	0.49	0.0045	0.48	0.50	0.67	0.0041	0.66	0.68	< 0.001	0.161	0.157	0.15	0.17
Error types													
Difference	0.01	0.0009	0.01	0.01	0.02	0.0012	0.02	0.02	< 0.001	0.067	0.242	0.16	0.33
Repetition	0.28	0.0029	0.27	0.28	0.27	0.0030	0.26	0.27	0.067	0.043	0.013	0.01	0.03
Wrong Principle	0.07	0.0017	0.06	0.07	0.06	0.0016	0.06	0.07	0.268	0.114	0.019	−0.01	0.05
Incomplete Correlate	0.09	0.0017	0.09	0.10	0.09	0.0018	0.09	0.10	0.778	0.085	0.004	−0.01	0.02
Indigenous													
Total Score RCPMT	0.45	0.0346	0.39	0.52	0.54	0.0606	0.42	0.66	0.230	0.198	0.085	−0.03	0.20
Error types													
Difference	0.02	0.0099	0.01	0.04	0.04	0.0150	0.01	0.07	0.195	0.023	0.331	−0.07	0.73
Repetition	0.32	0.0262	0.26	0.37	0.31	0.0304	0.25	0.37	0.809	0.128	0.014	−0.05	0.08
Wrong Principle	0.09	0.0156	0.06	0.13	0.09	0.0153	0.06	0.12	0.765	0.094	0.030	−0.10	0.16
Incomplete Correlate	0.12	0.0135	0.10	0.15	0.13	0.0218	0.09	0.17	0.842	0.078	0.020	−0.10	0.14

Note: M = arithmetic mean. SE = standard error. ICi = lower confidence interval. ICs = upper confidence interval. B = Bilingual. S = monolingual in Spanish. I = monolingual in an indigenous language. Total Score RCPMT = number of correct answers on the Raven's test; Difference = frequency of difference errors; Incomplete Correlate = frequency of errors due to incomplete correlate; Repetition = frequency of repetition errors; Wrong Principle = frequency of errors due to incorrect individualization. The table shows the p-values of the joint equality test in Eq. (2) and the RMSD in Eq. (3) of the methodological appendix, with the corresponding 95% confidence interval based on a "bootstrap" of 1000 replicates and the estimates and confidence intervals of βd for each group. Source: Prepared by the authors.

bilingual and monolingual children, a comparison is available with respect to poverty, but not with respect to overall Raven's test performance (Myers & Goldstein, 1979). It should be mentioned, however, that our results are in line with those obtained by Myers & Goldstein from a sample of lower-class monolingual English-speaking children and bilingual (English and Spanish-speaking) children enrolled in kindergarten, third and sixth grade of primary school. The authors reported that, although the bilingual children's verbal ability in both languages was lower than that of the monolinguals, the mean overall score and the error patterns of both groups on the Raven's test were equivalent. They concluded that, despite the apparent language differences, bilingual and monolingual children have comparable non-verbal reasoning cognitive abilities.

In addition, if we consider the distinction between cognitive functions of representation (abstract reasoning) and control (selective attention and interference suppression), the debate remains open as to the supposed advantages found in children with typical development on control-based tasks, but not in representation tasks (Bialystok, 2001). Following this line of argument, prior works have suggested that bilingualism in typically developing children could affect cognitive functions such as the selective attention to relevant aspects of a problem, the inhibition of distracting information, and the alternation between competing responses, which may translate to better performance as compared to their monolingual peers on the Ravens' tests (Craik & Bialystok, 2006). However, it cannot be ruled out that these differences will not translate into qualitatively different task resolution strategies or, at least, strategies that are different enough to generate differences in error patterns in the Raven's CPMT.

As for the results obtained when considering error patterns during primary education (from 5 to 12 years of age), Hypothesis 3 is supported for each of the three linguistic groups, with the results being consistent with those from previous studies considering typically developed children revealing an error pattern in which the most common error was R and the least common was D, while IC and WP had a similar frequency (Gunn & Jarrold, 2004; Petretto et al., 2021; Van Herwegen et al., 2011). Furthermore, the relative frequency of three error types (R, WP, and IC) is similar in children aged 5 to 8 and in those aged 9 to 12, regardless of the linguistic group. However, considering the two age cohorts, the results reveal how the intragroup differences in each linguistic group, are only statistically significant in the case of errors type D, in the group of monolingual children in Spanish.

According to observations from some standardization studies (Raven et al., 1990), the distribution of errors showed that the R error type is the most common, followed by IC, WP, and, finally, D errors. However, although Raven et al. (1990) observed that, when comparing error patterns of children having a mean age of 6.5 and 10.5 years, D errors go from 7 to 1%, WP errors go from 18 to 10%, R from 56 to 74% and IC from 19 to 15%. In fact, unlike Raven et al. (1990), we did not find a progressive increase in R errors.

Finally, from a methodological point of view, the results obtained are encouraging as they reveal how the method used, in addition to providing a greater level of comparison between experimental groups for a representative sample at the population level, also manages to do so without being excessively sensitive to the model used to estimate the GPS, or to the estimators used, and without omitting linguistic groups with less population representation from the analysis.

However, caution should be used when generalizing from the results since, because although the role of strong ignorability is clear in theory, in practice there is usually no way to prove the absence of differential selection on unobserved covariates. In this sense, although we have achieved a considerable balance between the three language groups (and the estimated GPS improve this balance in our entire sample) on

Table 10
Wald test: third estimator IPW.

Outcome	Bilingual (B) M	SE	ICi	ICs	Spanish (S) M	SE	ICi	ICs	Indigenous (I) M	SE	ICi	ICs	*p* value	Effect size	RMSD	ICi	ICs	Intergroup Comparisons *p* B-S	S-I	B-I
Entire sample (5–12 years of age)																				
Total Score RCPMT	0.53	0.0112	0.51	0.55	0.57	0.0079	0.56	0.59	0.53	0.0340	0.47	0.60	0.930	0.252	0.037	0.01	0.07	< 0.001	0.187	0.928
Error types																				
Difference	0.02	0.0026	0.02	0.03	0.02	0.0018	0.02	0.02	0.02	0.0070	0.01	0.03	0.910	0.069	0.019	−0.15	0.19	0.717	0.990	0.906
Repetition	0.29	0.0097	0.27	0.31	0.29	0.0085	0.28	0.31	0.29	0.0179	0.26	0.33	0.720	0.112	0.009	−0.02	0.03	0.310	0.965	0.722
Wrong Principle	0.08	0.0045	0.07	0.09	0.07	0.0037	0.07	0.08	0.07	0.0085	0.06	0.09	0.560	0.143	0.030	−0.04	0.10	0.306	0.811	0.565
Incomplete Correlate	0.10	0.0042	0.09	0.11	0.10	0.0034	0.10	0.11	0.11	0.0107	0.09	0.13	0.700	0.107	0.020	−0.03	0.08	0.838	0.637	0.702
Children between 9 and 12																				
Total Score RCPMT	0.63	0.0162	0.60	0.67	0.67	0.0098	0.65	0.69	0.61	0.0510	0.51	0.71	0.630	0.240	0.041	0.01	0.09	0.004	0.176	0.627
Error types																				
Difference	0.01	0.0031	0.01	0.02	0.01	0.0021	0.01	0.02	0.02	0.0077	0.01	0.03	0.850	0.119	0.063	−0.28	0.41	0.843	0.794	0.851
Repetition	0.23	0.0155	0.20	0.26	0.24	0.0141	0.21	0.27	0.24	0.0265	0.19	0.30	0.620	0.119	0.018	−0.03	0.06	0.247	0.890	0.617
Wrong Principle	0.05	0.0053	0.04	0.06	0.05	0.0047	0.04	0.06	0.05	0.0088	0.03	0.06	0.670	0.116	0.036	−0.09	0.16	0.446	0.893	0.674
Incomplete Correlate	0.08	0.0062	0.07	0.10	0.08	0.0049	0.07	0.09	0.08	0.0127	0.06	0.11	0.980	0.138	0.020	−0.07	0.11	0.475	0.774	0.984
Children between 5 and 8																				
Total Score RCPMT	0.44	0.0151	0.41	0.47	0.48	0.0065	0.47	0.49	0.47	0.0366	0.40	0.54	0.440	0.070	0.037	0.01	0.07	0.007	0.759	0.437
Error types																				
Difference	0.03	0.0045	0.02	0.04	0.03	0.0026	0.02	0.03	0.03	0.0111	0.01	0.05	0.770	0.066	0.060	−0.16	0.28	0.428	0.967	0.766
Repetition	0.33	0.0098	0.31	0.35	0.34	0.0071	0.32	0.35	0.33	0.0256	0.28	0.38	0.900	0.011	0.007	−0.03	0.05	0.478	0.933	0.896
Wrong Principle	0.10	0.0063	0.09	0.11	0.10	0.0048	0.09	0.11	0.09	0.0150	0.07	0.12	0.730	0.066	0.024	−0.06	0.11	0.409	0.938	0.729
Incomplete Correlate	0.12	0.0050	0.11	0.13	0.12	0.0037	0.11	0.13	0.12	0.0147	0.09	0.15	0.760	0.032	0.017	−0.05	0.09	0.572	0.866	0.757

Note: M = arithmetic mean. SE = standard error. ICi = lower confidence interval. ICs = upper confidence interval. B = Bilingual. S = monolingual in Spanish. I = monolingual in an indigenous language. Total Score RCPMT = number of correct answers on the Raven's test; Difference = frequency of difference errors; Incomplete Correlate = frequency of errors due to incomplete correlate; Repetition = frequency of repetition errors; Wrong Principle = frequency of errors due to incorrect individualization. The table shows the p-values of the joint equality test in Eq. (2) and the RMSD in Eq. (3) of the methodological appendix, with the corresponding 95% confidence interval based on a "bootstrap" of 1000 replicates and the estimates and confidence intervals of βd for each group. Source: Prepared by the authors.

Table 11
Wald test: third estimator IPW.

	Children between 5 and 8				Children between 9 and 12								
Variable	M	SE	ICi	ICs	M	SE	ICi	ICs	*p* value	Effect size	RMSD	ICi	ICs
All													
Total Score RCPMT	0.48	0.0042	0.47	0.49	0.66	0.0042	0.65	0.67	< 0.001	0.188	0.158	0.15	0.17
Error types													
Difference	0.01	0.0009	0.01	0.02	0.02	0.0012	0.02	0.02	< 0.001	0.023	0.234	0.16	0.31
Repetition	0.28	0.0027	0.27	0.29	0.27	0.0030	0.27	0.28	0.036	0.121	0.013	0.01	0.03
Wrong Principle	0.07	0.0015	0.07	0.07	0.07	0.0016	0.06	0.07	0.212	0.088	0.019	−0.01	0.04
Incomplete Correlate	0.10	0.0014	0.09	0.1	0.10	0.0018	0.09	0.10	0.968	0.070	< 0.001	−0.01	0.02
Bilingual													
Total Score RCPMT	0.44	0.0121	0.42	0.47	0.63	0.0125	0.61	0.65	< 0.001	0.182	0.172	0.14	0.20
Error types													
Difference	0.02	0.0028	0.01	0.02	0.03	0.0034	0.02	0.03	0.056	0.050	0.176	0.01	0.35
Repetition	0.30	0.0071	0.28	0.31	0.29	0.0084	0.27	0.30	0.366	0.116	0.015	−0.01	0.04
Wrong Principle	0.08	0.0050	0.07	0.09	0.08	0.0048	0.07	0.09	0.612	0.090	0.021	−0.03	0.07
Incomplete Correlate	0.10	0.0042	0.09	0.11	0.11	0.0052	0.10	0.12	0.621	0.061	0.015	−0.03	0.06
Spanish													
Total Score RCPMT	0.49	0.0046	0.48	0.5	0.67	0.0043	0.66	0.68	< 0.001	0.147	0.156	0.15	0.17
Error types													
Difference	0.01	0.0009	0.01	0.01	0.02	0.0012	0.02	0.02	< 0.001	0.153	0.232	0.15	0.32
Repetition	0.28	0.0029	0.27	0.28	0.27	0.0031	0.26	0.28	0.083	0.043	0.012	0.01	0.02
Wrong Principle	0.07	0.0017	0.06	0.07	0.06	0.0016	0.06	0.07	0.267	0.150	0.019	−0.01	0.05
Incomplete Correlate	0.10	0.0017	0.09	0.1	0.09	0.0020	0.09	0.10	0.852	0.067	0.003	−0.01	0.02
Indigenous													
Total Score RCPMT	0.46	0.0377	0.39	0.53	0.54	0.0601	0.42	0.65	0.274	0.190	0.078	−0.04	0.19
Error types													
Difference	0.02	0.0116	0.01	0.04	0.05	0.0168	0.01	0.08	0.125	0.021	0.397	−0.04	0.83
Repetition	0.30	0.0285	0.25	0.36	0.29	0.0317	0.23	0.35	0.704	0.123	0.026	−0.05	0.11
Wrong Principle	0.09	0.0158	0.06	0.12	0.08	0.0156	0.05	0.11	0.618	0.089	0.057	−0.10	0.22
Incomplete Correlate	0.12	0.0147	0.09	0.15	0.13	0.0227	0.09	0.17	0.725	0.073	0.036	−0.09	0.17

Note: M = arithmetic mean. SE = standard error. ICi = lower confidence interval. ICs = upper confidence interval. B = Bilingual. S = monolingual in Spanish. I = monolingual in an indigenous language. Total Score RCPMT = number of correct answers on the Raven's test; Difference = frequency of difference errors; Incomplete Correlate = frequency of errors due to incomplete correlate; Repetition = frequency of repetition errors; Wrong Principle = frequency of errors due to incorrect individualization. The table shows the p-values of the joint equality test in Eq. (2) and the RMSD in Eq. (3) of the methodological appendix, with the corresponding 95% confidence interval based on a "bootstrap" of 1000 replicates and the estimates and confidence intervals of βd for each group. Source: Prepared by the authors.

our observed covariates, it does not necessarily make the groups equivalent on unobserved but important covariates (Steiner et al., 2010). As part of this discussion three additional considerations should be made: (1) due to the use of the probability sampling method, the sample size was only 41 participants in the case of monolinguals belonging to an ethnic community using an indigenous language; (2) matching our four variables may not eliminate their causal impact if they have indirect causes (e.g., interaction effects) on the dependent variables. For example, interactions between age and poverty (Lipina, 2016; Najman et al., 2009; Strauß, Venables, & Zentner, 2023), age and sex (Lynn, Allik, & Irwing, 2004a) or age and linguistic group (Bialystok, Craik, Klein, & Viswanathan, 2004); and (3) uninvestigated variables like nutritional status of children (Akubuilo et al., 2020), mother-child relationship (McGowan & Johnson, 1984), maternal cognitive skills (Gredebäck et al., 2022; Wachs et al., 2007), and other antecedents of childhood intelligence, could also cause the differences found in this study.

Without overlooking the limitations derived from the methodological approach, and from the restrictions imposed by the available data and variables, this line of study may be quite fruitful and would benefit from future examination of other learning contexts, as well as the use of other complementary theoretical approaches to analyze the error types considered in this work. In fact, it leaves room to advance this line of research with complementary views to examine error types, for example, examining item level to determine its degree of difficulty (Facon & Nuchadee, 2010; Van Herwegen et al., 2011), typology (Lynn, Allik, & Irwing, 2004b), reaction time for its resolution (Soulières et al., 2009), or even its comparison with computational models (Carpenter, Just, & Shell, 1990; Kunda et al., 2016) and Artificial Intelligence (Hua & Kunda, 2020; Raudies & Hasselmo, 2017; Yang, McGreggor, & Kunda, 2022). It would also be interesting to introduce additional covariates on individual characteristics, cultural background, or language skills (context of language use, history of acquisition, self-reported proficiency, etc.), given that the experience of bilingualism itself is dynamic and consists of multiple related dimensions (Luk & Bialystok, 2013). Future research should focus on the role of language competence of monolinguals and bilinguals, given that, in our study, we have been unable to do this, since data of this nature were not available. As for minority languages, Gonthier (2022) reviewed the literature on cultural differences in the processing of visuo-spatial materials, identifying some sources of differences between ethnic groups that could help explain differences in mean scores on the specific context of reasoning tests such as the Raven's CPMT. Among those sources, the cognitive representation of color, space, relationships between objects, and numerosity may be influenced by the use of specific verbal coding in some minority languages (e.g., Mayan languages spoken in Mexico, such as Tzeltal), which could be explored as possible factors influencing the mental representation of abstract relations between objects in a visual display of the Raven's task.

We also note that, in addition to the contributions mentioned, this work has certain practical implications in the educational field. It provides new evidence to better understand the role of environmental factors in determining non-verbal abstract reasoning in bilingual or monolingual children from multiethnic communities mediated by poverty and with linguistic diversity. We focused on a multi-ethnic impoverished environment in a Latin-American region (Mexico), containing minorities recognized by the Global Education Agenda 2030 as vulnerable groups affected by poverty and inequality (UNESCO, 2020). It is reasonable to assume that the identification of risk factors to which children are exposed may optimize the design of educational interventions in conditions that are initially more unfavorable for modulating the degree of their impact on children's reasoning abilities and academic achievement. For example, a longitudinal study by Woumans et al. (2016) suggests that, after one year of attending a bilingual immersion schooling program, five-year-olds significantly improved their general cognitive abilities, as assessed by intelligence, as compared to their kindergarten peers attending a monolingual program.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank Dr. Richard Haier and the anonymous reviewers for contributing to this article with their valuable comments. We would also like to thank Dr. M. Fernandez Cahill for proofreading this manuscript, and the Mexican Family Life Survey (MxFLS) for providing the database used in this study. This work was funded by the Ministry of Science and Innovation of Spain [PID2019-110476RB-100].

Appendix A. Methodological Appendix

Each unit i in our sample, $i = 1, 2, \ldots, N$, comes from one of the possible k groups, with $D_i \in \{1, 2, \ldots, k\}$ denoting the location of individual i. For our purposes, it is convenient to write the possible outcomes of interest as $Y_i(0, d)$, where d denotes the group and 0 emphasizes the fact that we focus exclusively on the control groups. Therefore, $Y_i(0, d)$ is the result (for example, linguistic characteristic, such as bilingualism, monolingualism, etc.) that the individual would obtain if that person were a control at site d. The data we observe for each unit are (Y_i, D_i, X_i), where X_i is a set of pretreatment covariates and $Y_i = Y(0, D_i)$.

S is a subset of the support of X, χ. In this study, our parameters of interest are:

$$\beta_d \equiv \beta_d(S) = E[Y(0,d)\,|X \in S] \text{ for } d = 1, 2, \ldots, k \tag{1}$$

Region S may include all or part of the support for X and is determined by the region of overlap, as discussed below. To simplify the notation, we omit the conditioning on $X \in S$ unless it is necessary for clarity. The object in Eq. (1) gives the average potential outcome under the control treatment at location d for someone with $X \in S$ randomly selected from any of the k groups. This differs from $E[Y(0,d)|D = d, X \in S]$, which yields the mean control result in group d for someone with $X \in S$ randomly selected from group d. While the latter quantity is identified from the data due to random assignment of treatment within group d, β_d is not.

Instead of focusing on Eq. (1), we could have focused on $E[Y(0,d)|D = f, X \in S_f]$ for $d = 1, 2, \ldots, k$, which yields the mean result for individuals in a particular group $f \in \{1, 2, \ldots, k\}$ within the region of overlap S_f for group f. By concentrating on the mean of all k groups in Eq. (1), we prevent our results from depending on the selection of group f as the reference point. However, the estimation problem becomes increasingly challenging given the need to find comparable individuals in each of the groups for each individual in groups k (instead of finding comparable individuals in each site only for those in group f).

We evaluate the performance of the strategies presented below in several ways. First, given estimates $\widehat{\beta_1}, \ldots, \widehat{\beta_k}$ of the corresponding parameters in Eq. (1), we test the joint hypothesis,

$$\beta_1 = \beta_2 = \ldots = \beta_k \tag{2}$$

One drawback of this approach is that the null hypothesis in Eq. (2) may not be rejected, simply because the variance of the estimators is high and not because all the estimated values of $\widehat{\beta_1}$'s are sufficiently close to each other. Therefore, we focus our analysis primarily on a general measure of distance between the estimated means. By allowing $\widehat{\mu} = k^{-1}\sum_{d=1}^{k}\widehat{\beta_1}$, the root mean square distance is defined as:

$$\text{RMSD} = \frac{1}{|\widehat{\mu}|}\sqrt{\frac{1}{k}\sum\nolimits_{d=1}^{k}\left(\widehat{\beta_d} - \widehat{\mu}\right)^2} \tag{3}$$

where we divide by $|\widehat{\mu}|$ to facilitate its interpretation and comparison between results. Due to pure sample variation, we would never expect RMSD to be equal to 0, even if D_i were randomly assigned.

References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207–245. https://doi.org/10.3102/0034654310368803

Akubuilo, U. C., Iloh, K. K., Onu, J. U., Iloh, O. N., Ubesie, A. C., & Ikefuna, A. N. (2020). Nutritional status of primary school children: Association with intelligence quotient and academic performance. *Clinical Nutrition ESPEN, 40*, 208–213. https://doi.org/10.1016/j.clnesp.2020.09.019

Antoniou, M. (2019). The advantages of bilingualism debate. Annual Review of Linguistics, 5, 395–415. doi:10.1146/annurev-linguistics-011718-011820.

Anum, A. (2022). Does socio-economic status have different impact on fluid and crystallized abilities? Comparing scores on Raven's progressive matrices, Kaufman assessment battery for children II story completion and Kilifi naming test among children in Ghana. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.880005

Arán-Filippetti, V., Krumm, G. L., & Raimondi, W. (2015). Funciones ejecutivas y sus correlatos con inteligencia cristalizada y fluida: Un estudio en niños y adolescentes. *Revista Neuropsicología Latinoamericana, 7*(2), 24–33. https://doi.org/10.5579/rnl.2015.0213

Ayoub, C., O'Connor, E., Rappolt-Schlictmann, G., Vallotton, C., Raikes, H., & Chazan-Cohen, R. (2009). Cognitive skill performance among young children living in poverty: Risk, change, and the promotive effects of early head start. *Early Childhood Research Quarterly, 24*(3), 289–305. https://doi.org/10.1016/j.ecresq.2009.04.001

Ben-Zeev, S. (1977a). The influence of bilingualism on cognitive strategy and cognitive development. *Child Development, 3*(48), 1009–1018. https://doi.org/10.2307/1128353

Ben-Zeev, S. (1977b). The effect of bilingualism in children from spanish-english low economic neighborhoods on cognitive development and cognitive strategy. *Working Papers on Bilingualism, 14*.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin, 143*(3), 233–262. https://doi.org/10.1037/bul0000099 (PubMed: 28230411).

Bialystok, E., Craik, F. I., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest, 10*(3), 89–129. https://doi.org/10.1177/1529100610387084

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*(2), 290. https://psycnet.apa.org/doi/10.1037/0882-7974.19.2.290.

Bialystok, E., & Shorbagi, S. H. (2021). Subtle increments in socioeconomic status and bilingualism jointly affect children's verbal and nonverbal performance. *Journal of Cognition and Development, 22*(3), 467–490. https://doi.org/10.1080/15248372.2021.1901711

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review, 97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1–22. https://doi.org/10.1037/h0046743

Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences, 10*(3), 131–138. https://doi.org/10.1016/j.tics.2006.01.007

Crosnoe, R., Leventhal, T., Wirth, R. J., Pierce, K. M., Pianta, R. C., & NICHD Early Child Care Research Network. (2010). Family socioeconomic status and consistent environmental stimulation in early childhood. *Child Development, 81*(3), 972–987. https://doi.org/10.1111/j.1467-8624.2010.01446.x

Daniele, V. (2021). Socioeconomic inequality and regional disparities in educational achievement: The role of relative poverty. *Intelligence, 84*, 101515. https://doi.org/10.1016/j.intell.2020.101515

De Bruin, A., Dick, A. S., & Carreiras, M. (2021). Clear theories are needed to interpret differences: Perspectives on the bilingual advantage debate. *Neurobiology of Language, 2*(4), 433–451. https://doi.org/10.1162/nol_a_00038

De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science, 26*(1), 99–107. https://doi.org/10.1177/0956797614557866

Deary, I., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews. Neuroscience, 11*, 201–211. https://doi.org/10.1038/nrn2793

Donnelly, S., Brooks, P. J., & Homer, B. D. (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychonomic Bulletin & Review, 26*(4), 1122–1147. https://doi.org/10.3758/s13423-019-01567-z

Engel de Abreu, P. M., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor: Enhanced cognitive control in low-income minority children. *Psychological Science, 23*(11), 1364–1371. https://doi.org/10.1177/0956797612443836

Facon, B., Magis, D., Nuchadee, M. L., & De Boeck, P. (2011). Do Raven's colored progressive matrices function in the same way in typical and clinical populations? Insights from the intellectual disability field. *Intelligence, 39*(5), 281–291. https://doi.org/10.1016/j.intell.2011.04.002

Facon, B., & Nuchadee, M. L. (2010). An item analysis of Raven's colored progressive matrices among participants with down syndrome. *Research in Developmental Disabilities, 31*(1), 243–249. https://doi.org/10.1016/j.ridd.2009.09.011

Fajgelj, S., Bala, G., & Katić, R. (2010). Latent structure of Raven's colored progressive matrices. *Collegium antropologicum, 34*(3), 1015–1026.

Farran, E. K., Atkinson, L., & Broadbent, H. (2016). Impaired spatial category representations in Williams syndrome; an investigation of the mechanistic contributions of non-verbal cognition and spatial language performance. *Frontiers in Psychology, 1868*(7), 1–12. https://doi.org/10.3389/fpsyg.2016.01868

Feres, J. C., & Mancero, X. (2001). Enfoques para la medición de la pobreza: breve revisión de la literatura. In *, 4. CEPAL Serie Estudios Estadísticos y Prospectivos*. https://hdl.handle.net/11362/4740.

Flores, C. A., & Mitnik, O. A. (2013). Comparing treatments across labour markets: An assessment of nonexperimental multiple-treatment strategies. *Review of Economics and Statistics, 95*(5), 1691–1707. https://doi.org/10.1162/REST_a_00373

Goharpey, N., Crewther, D. P., & Crewther, S. G. (2013). Problem solving ability in children with intellectual disability as measured by the Raven's colored progressive matrices. *Research in Developmental Disabilities, 34*(12), 4366–4374. https://doi.org/10.1016/j.ridd.2013.09.013

Gonthier, C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: An integrative review and a model for matrices tasks. *Cognitive Research: Principles and Implications, 7*(1), 1–27. https://doi.org/10.1186/s41235-021-00350-w

Gredebäck, G., Hall, J., & Lindskog, M. (2022). Fluid intelligence in refugee children. A cross-sectional study of potential risk and resilience factors among Syrian refugee children and their parents. *Intelligence, 94*, 101684. https://doi.org/10.1016/j.intell.2022.101684

Grosjean, F., & Li, P. (2013). *The psycholinguistics of bilingualism*. Wiley-Blackwell.

Gunn, D. M., & Jarrold, C. (2004). Raven's matrices performance in down syndrome: Evidence of unusual errors. *Research in Developmental Disabilities, 25*(5), 443–457. https://doi.org/10.1016/j.ridd.2003.07.004

Gunnerud, H. L., Ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin, 146*(12), 1059–1083. https://doi.org/10.1037/bul0000301

Gutiérrez, J. P., Heredia-Pi, I., Hernández-Serrato, M. I., Pelcastre-Villafuerte, B. E., Torres-Pereda, P., & Reyes-Morales, H. (2020). Desigualdades en el acceso a servicios, base de las políticas para la reducción de la brecha en salud. *Salud Pública de México, 61*(6), 726–733. https://doi.org/10.21149/10561

Haier, R. J. (2016). *The neuroscience of intelligence*. Cambridge University Press.

Hernán, M. A., & Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health, 60*(7), 578–586. https://doi.org/10.1136/jech.2004.029496

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments, *226164*. *Applied Bayesian modeling and causal inference from incomplete-data perspectives* (pp. 73–84).

Hua, T., & Kunda, M. (2020). Modeling gestalt visual reasoning on Raven's Progressive Matrices using generative image inpainting techniques. In *Proceedings of the Eighth Annual Conference on Advances in Cognitive Systems (ACS)*.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika, 87*(3). https://doi.org/10.1093/biomet/87.3.706

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5

INEGI. (2021). *Mexico: Encuesta Intercensal 2015*. https://www.inegi.org.mx/rnm/index.php/catalog/214/pdf-documentation.

Jacobs Paul, I., & Mary, Vandeventer. (1970). Information in wrong responses. *Psychological Reports, 26*(1), 311–315. https://doi.org/10.2466/pr0.1970.26.1.311

Johann, V., Enke, S., Gunzenhause, C., Könen, T., Saalbach, H., & Karbach, J. (2022). Executive functions in mono- and bilingual children: Factor structure and relations with fluid intelligence. *Journal of Experimental Child Psychology, 224*. https://doi.org/10.1016/j.jecp.2022.105515

Kalashnikova, M., Pejovic, J., & Carreiras, M. (2021). The effects of bilingualism on attentional processes in the first year of life. *Developmental Science, 24*(2), Article e13011. https://doi.org/10.1111/desc.13011

Kaniel, S., & Fisherman, S. (1991). Level of performance and distribution of errors in the progressive matrices test: A comparison of Ethiopian immigrant and native Israeli adolescents. *International Journal of Psychology, 26*(1), 25–33. https://doi.org/10.1080/00207599108246847

Kaushanskaya, M., & Prior, A. (2015). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism: Language and Cognition, 18*(1), 27–28. https://doi.org/10.1017/S1366728914000510

Kaya, F., Stough, L. M., & Juntune, J. (2016). The effect of poverty on the verbal scores of gifted students. *Educational Studies, 42*(1), 85–97. https://doi.org/10.1080/0305569.2016.1148585

Knauer, H. A., Ozer, E. J., Dow, W., & Fernald, L. C. H. (2018). Stimulating parenting practices in indigenous and non-indigenous Mexican communities. *International Journal of Environmental Research and Public Health, 15*(1), 29. https://www.mdpi.com/1660-4601/15/1/29.

Kroll, J. F., & Bialystok, E. (2013a). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology, 25*(5), 497–514. https://doi.org/10.1080/20445911.2013.799170

Kroll, J. F., & Bialystok, E. (2013b). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology (Hove, England), 25*(5). https://doi.org/10.1080/20445911.2013.799170

Kunda, M., Soulières, I., Rozga, A., & Goel, A. (2013). Methods for classifying errors on the Raven's standard progressive matrices test. *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*(35), 2796–2801.

Kunda, M., Soulières, I., Rozga, A., & Goel, A. (2016). Error patterns on the Raven's standard progressive matrices test. *Intelligence, 59*, 181–198. https://doi.org/10.1016/j.intell.2016.09.004

Laborda, L., Elosúa, M. R., & Gómez-Veiga, I. (2019). Ethnicity and intelligence in children exposed to poverty environments: An analysis using the Oaxaca-blinder decomposition. *Intelligence, 72*, 49–58. https://doi.org/10.1016/j.intell.2018.11.008

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin, 144*(4), 394–425. https://doi.org/10.1037/bul0000142

Lipina, S. (2016). Critical considerations about the use of poverty measures in the study of cognitive development. *International Journal of Psychology*. https://doi.org/10.1002/ijop.12282

Lipina, S., & Evers, K. (2017). Neuroscience of childhood poverty: Evidence of impacts and mechanisms as vehicles of dialog with ethics. *Frontiers in Psychology, 8*(61). https://doi.org/10.3389/fpsyg.2017.00061

Lowe, C. J., Cho, I., Goldsmith, S. F., & Morton, J. B. (2021). The bilingual advantage in children's executive functioning is not related to language status: A meta-analytic review. *Psychological science, 32*(7), 1115–1146. https://doi.org/10.1177/0956797621993108

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Lynn, R., Allik, J., & Irwing, P. (2004a). Sex differences on three factors identified in Raven's standard progressive matrices. *Intelligence, 32*(4), 411–424. https://doi.org/10.1016/j.intell.2004.06.007

Lynn, R., Allik, J., & Irwing, P. (2004b). Sex differences on three factors identified in Raven's standard progressive matrices. *Intelligence, 32*(4), 411–424. https://doi.org/10.1016/j.intell.2004.06.007

Lynn, R., Fuerst, J., & Kirkegaard, E. O. W. (2018). Regional differences in intelligence in 22 countries and their economic, social and demographic correlates: A review. *Intelligence, 69*, 24–36. https://doi.org/10.1016/j.intell.2018.04.004

Matzen, L. B. V., Van der Molen, M. W., & Dudink, A. C. (1994). Error analysis of Raven test performance. *Personality and Individual Differences, 16*(3), 433–445. https://doi.org/10.1016/0191-8869(94)90070-1

McGowan, R. J., & Johnson, D. L. (1984). The mother-child relationship and other antecedents of childhood intelligence: A causal analysis. *Child Development*, 810–820. https://psycnet.apa.org/doi/10.2307/1130132.

Millones, D. M., Flores-Mendoza, C., & Rivalles, R. M. (2015). Intelligence in Peru: Students' results in Raven and its relationship to SES. *Intelligence, 51*, 71–78. https://doi.org/10.1016/j.intell.2015.05.004

Myers, B., & Goldstein, D. (1979). Cognitive development in bilingual and monolingual lower-class children. *Psychology in the Schools, 16*(1), 137–142. https://doi.org/10.1002/1520-6807(197901)16:1%3C137::AID-PITS2310160124%3E3.0.CO;2-Z

Najman, J. M., Hayatbakhsh, M. R., Heron, M. A., Bor, W., O'Callaghan, M. J., & Williams, G. M. (2009). The impact of episodic and chronic poverty on child cognitive development. *The Journal of Pediatrics, 154*(2), 284–289. https://doi.org/10.1016/j.jpeds.2008.08.052

Nistal, M. T. F. (2014). Datos normativos de las Matrices Progresivas Coloreadas en niños indígenas yaquis. *Anuario de Psicología, 44*(3), 373–385. https://raco.cat/index.php/AnuarioPsicologia/article/view/285849.

Oxfam. Por mi raza hablará la desigualdad. Efectos de las características étnico-raciales en la desigualdad de oportunidades en México. https://oxfammexico.org/por-mi-raza-hablara-la-desigualdad-resumen-ejecutivo/.

Paap, K. (2019). The bilingual advantage debate: Quantity and quality of the evidence. In J. W. Schwieter (Ed.), *The handbook of the neuroscience of multilingualism* (pp. 701–735). Wiley-Blackwell. https://doi.org/10.1002/9781119387725.ch34.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex, 69*, 265–278. https://doi.org/10.1016/j.cortex.2015.04.014

Pesta, B. J., Kirkegaard, E. O. W., Nijenhuis, J., Lasker, J., & Fuerst, J. G. R. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence, 78*, Article 101408. https://doi.org/10.1016/j.intell.2019.101408

Petretto, D. R., Grassi, P., Masala, C., & Nicotra, E. F. (2021). Pattern of errors in Raven's colored progressive matrices and their use in the clinical assessment of intelligence. In *IEEE international symposium on medical measurements and applications (MeMeA)* (pp. 1–6). https://doi.org/10.1109/MeMeA52024.2021.9478720

Platt, J. M., McLaughlin, K., Luedtke, A., Ahern, J., Kaufman, A., & Keyes, K. (2018). Targeted estimation of the relationship between childhood adversity and fluid intelligence in a US population sample of adolescent. *American Journal of Epidemiology, 187*, 7. https://doi.org/10.1093/aje/kwy006

Rahu, K., Rahu, M., Pullmann, H., & Allik, J. (2010). Effect of birth weight, maternal education and prenatal smoking on offspring intelligence at school age. *Early Human Development, 86*(8), 493–497. https://doi.org/10.1016/j.earlhumdev.2010.06.010

Raudies, F., & Hasselmo, M. E. (2017). A model of symbolic processing in Raven's progressive matrices. *Biologically Inspired Cognitive Architectures, 21*, 47–58. https://doi.org/10.1016/j.bica.2017.07.003

Raven, J. C., Court, J. H., & Raven, J. (1990). Section 2: Coloured progressive matrices (1990 edition, with US norms). In *Manual for the Raven's progressive matrices and vocabulary scales*. Oxford: Oxford Psychologist Press.

Raven, J. C., Court, J. H., & Raven, J. (2006). *Test de Matrices Progresivas*. Escalas Coloreada. General y Avanzada. Paidos Iberica Eds.

Rubalcava, L., & Teruel, G. (2013). Mexican family life survey, Third Wave. http://www.ennvih-mxfls.org/english/assets/usersguidev2-(1).pdf.

Rubin, D. B. (1990). Formal mode of statistical inference for causal effects. *Journal of statistical planning and inference, 25*(3), 279–292. https://doi.org/10.1016/0378-3758(90)90077-8

Ryan, R. M., Fauth, R. C., & Brooks-Gunn, J. (2006). Childhood poverty: Implications for school readiness and early childhood education. In B. Spodek, & O. N. Saracho (Eds.), *Handbook of research on the education of young children* (pp. 323–346). Lawrence Erlbaum Associates Publishers.

Segretin, M. S., Hermida, M. J., Prats, L. M., Fracchia, C. S., Ruetti, E., & Lipina, S. J. (2016). Childhood poverty and cognitive development in Latin America in the 21st century. In D. D. Preiss (Ed.), *152. Child and adolescent development in Latin America* (pp. 9–29). New Directions for Child and Adolescent Development.

Sen, A. K. (2000). *What is development about? Frontiers of Development Economics: The future in perspective* (pp. 506–513).

Soulières, I., Dawson, M., Samson, F., Barbeau, E. B., Sahyoun, C. P., Strangman, G. E., … Mottron, L. (2009). Enhanced visual processing contributes to matrix reasoning in autism. *Human Brain Mapping, 30*(12), 4082–4107. https://doi.org/10.1002/hbm.20831

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods, 15*(3), 250. https://doi.org/10.1037/a0018719

Strauß, H., Venables, P., & Zentner, M. (2023). Associations between early childhood poverty and cognitive functioning throughout childhood and adolescence: A 14-year prospective longitudinal analysis of the Mauritius child health project. *PLoS One, 18*(2), Article e0278618. https://doi.org/10.1371/journal.pone.0278618

Tong, S., Baghurst, P., Vimpani, G., & McMichael, A. (2007). Socioeconomic position, maternal IQ, home environment, and cognitive development. *The Journal of Pediatrics, 151*(3), 284–288. https://doi.org/10.1016/j.jpeds.2007.03.020

Treccani, B., Argyri, E., Sorace, A., & Della Sala, S. (2009). Spatial negative priming in bilingualism. *Psychonomic Bulletin & Review, 16*(2), 320–327. https://doi.org/10.3758/PBR.16.2.320

UNESCO. (2020). *Global education monitoring report, 2020*. Inclusion and education: all means all. https://unesdoc.unesco.org/ark:/48223/pf0000373718.

UNICEF-CONEVAL. (2019). *Pobreza infantil y adolescente en México 2008–2016. Dónde vive y qué características tiene la población de 0 a 17 años en situación de pobreza*. UNICEF-CONEVAL.

Van de Vijver, F. J., & Brouwers, S. A. (2009). Schooling and basic aspects of intelligence: A natural quasi-experiment in Malawi. *Journal of Applied Developmental Psychology, 30*(2), 67–74. https://doi.org/10.1016/j.appdev.2008.10.010

Van den Noort, M., Struys, E., & Bosch, P. (2019). Individual variation and the bilingual advantage—Factors that modulate the effect of bilingualism on cognitive control and cognitive reserve. *Behavioral Science, 9*(12), 120. https://doi.org/10.3390/bs9120120 (PubMed: 31766485).

Van Herwegen, J., Farran, E., & Annaz, D. (2011). Item and error analysis on Raven's Coloured progressive matrices in Williams syndrome. *Research in Developmental Disabilities, 32*(1), 93–99. https://doi.org/10.1016/j.ridd.2010.09.005

Wachs, T. D., Chang, S. M., Walker, S. P., & Gardner, J. M. M. (2007). Relation of birth weight, maternal intelligence and mother–child interactions to cognitive and play competence of Jamaican two-year old children. *Intelligence, 35*(6), 605–622. https://doi.org/10.1016/j.intell.2006.11.005

Woumans, E., Surmont, J., Struys, E., & Duyck, W. (2016). The longitudinal effect of bilingual immersion schooling on cognitive control and intelligence. *Language Learning, 66*(S2), 76–91. https://doi.org/10.1111/lang.12171

Yang, Y., McGreggor, K., & Kunda, M. (2022). Visual-imagery-based analogical construction in geometric matrix reasoning Task. *arXiv preprint*. arXiv:2208.13841.