Cultural Bias in Intelligence Assessment Using a "Culture-Free" Test in Moroccan Children

Alvaro Lozano-Ruiz[1, 2], Ahmed F. Fasfous[1, 3], Inmaculada Ibanez-Casas[1, 4], Francisco Cruz-Quintana[1, 2], Miguel Perez-Garcia[1, 2], and María Nieves Pérez-Marfil[1, 2]

[1]Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain

[2]Department of Personality, Evaluation, and Psychological Treatment, University of Granada, Granada, Spain

[3]Department of Social Sciences, Bethlehem University, Bethlehem, Palestine

[4]Department of Psychology, State University of New York at Plattsburgh, New York, United States

Author Note

Alvaro Lozano-Ruiz https://orcid.org/0000-0003-1661-6638

Ahmed F. Fasfous https://orcid.org/0000-0002-4812-9853

Inmaculada Ibanez-Casas https://orcid.org/0000-0002-1007-8966

Francisco Cruz-Quintana https://orcid.org/0000-0001-9805-5939

Miguel Perez-Garcia https://orcid.org/0000-0003-4775-7556

María Nieves Pérez-Marfil https://orcid.org/0000-0001-6329-8770

Correspondence concerning this article should be addressed to Ahmed F. Fasfous, Department of Social Sciences, Bethlehem University, Bethlehem, Palestine.

Email: afasfous@bethlehem.edu

Abstract

Objective: Previous research has shown that cognitive tests can lead to misclassification when applying non-representative norms to measure cognitive performance.The objective of this study was to investigate whether this misclassification also occurs with anon-verbal so-called“culture-free” intelligence test administered to different age groups. Method: The intelligence of a sample of healthy Moroccan children (N = 147) ages 7, 9, and 11 was assessed using the Coloured Raven’s Progressive Matrices (CPM). Raw scores were used to study age differences, as well asmisclassificationswhenapplying the normsof three countries culturallydifferent from Morocco (United Kingdom, Spain, and Oman). Results: Intelligence performance was within the normal range when non-representative norms were applied to the Moroccan raw scores. Misclassifications accounted fora large percentage of the participants that supposedly displayed intelligence deficits, especially when applying the Britishnorms. Up to 15.68% of the healthychildrenfell within the “intellectually impaired” range, and up to 62.5% fell “below average”, with these percentages especially higher at older ages.Conclusions: Our findings confirm that “culture-free” tests should be adapted to each culture and applied together with their culture’s specific norms to prevent misclassification and allow for a better, unbiased neuropsychological assessment.

Keywords: Cultural Test Bias; Cultural Factors; Culture Fair Intelligence Test; Cross Cultural Differences; Raven Coloured Progressive Matrices

Cultural Bias in Intelligence Assessment Using a Culture-Free Test in Moroccan Children

Despite the wide availability of data from different countries, cultures, and contexts, the influence of culture factors on cognitive abilities and intelligence performance,is still unclear (Gerogas, 2003).This is especially apparent during child development stages (Fasfous et al., 2013). According to Greenfield (1997), psychological tests are based upon the values and behavioral expectations of western societies in which they were developed, thus do not easily cross cultures. To highlight this point and prevent misinterpretation of test scores, Fujii (2017) recommended that performances on neuropsychological tests by diverse examinees should be referenced with “on western tests”.

Sternberg (2004) argues that intelligent behavior is a culturally defined construct that is directly tied to addressing challenges in one’s environment. This contention is supported by studies with Kenyans (Sternberg et al., 2001) and Alaskan Natives (Grigorenko et al., 2004) demonstrating that measures evaluating cultural specific characteristics of intelligence were a better indicator of everyday functioning than western based cognitive tests.Moreover, performances on these tests were unrelated (Sternberg, 1990, 2007).

Numerous factors have been found to impact performance on western cognitive and intelligence tests which can account for cross cultural differences including socioeconomic status (SES) (von Stumm & Plomin, 2015; Shuttleworth-Edwards et al., 2014), and educational factors such as academic achievement (Deary et al., 2007; Strenze, 2007), parental education level and family social climate (Georgas, 2003), quality of education (Rindermann, 2008), and literacy (Ardila et al., 2010). The ubiquitous impact of such factors in addition to other social technological factors on test performances over time has been documented across the globe and referred to as the Flynn Effect (Flynn, 2012; Raven, 2000). Additionally, test results can be influenced by other subtle cultural factors, such as attitudes toward time (Agranovich et al., 2011), familiarity with previous psychological tests (Puente & Perez-Garcia, 2000), or urban/rural environments (Melikyan et al., 2020).

To reduce the impact of these cultural variables on intelligence assessment, “culture free” tests such as the Cattell Culture Fair Intelligence Test (Cattell, 1973), BETA-III (Kellogg & Morton, 1999), and Raven’s Progressive Matrices (Raven, 1998) were developed (Cattell, 1940; Cattell et al., 1941). These tests were deemed “culture free” primarily due to the minimization of language requirements. However, studies with samples from different non-English speaking countries have consistently reported poorer performances when using U.S. norms (Fasfous et al., 2013;Daugherty et al., 2017), suggesting that these nonverbal tests are influenced by cultural variables (Ardila, 2005; Rosselli & Ardila, 2003). Similar weaker performances have been found in different non-English speaking cultures whenapplying the normative standards of memory and executive function tasks in African Americans (Norman et al., 2011).

Raven’s Coloured Progressive Matrices (CPM) is one of the most widely used “culture-free” cognitive ability tests for children between the ages of 5 and 11 (Raven et al., 1998). Although the first normative data was collected in 1942, currently the most widely used norms come from a follow-up study conducted on 598 children from Dumfries, UK (1982). CPM has been used internationally, with norms from Spain (Raven et al., 2011), Australia (Cotton et al., 2005), South Africa (Linstrom et al., 2008), Brazil (Bandeira et al., 2004), and countries in the Arab world such as Oman (Kazem et al., 2009), United Arab Emirates (Khaleefa & Lynn, 2008), Libya (Lynn et al., 2008), Sudan (Bakhiet et al., 2017), and Egypt (Ziada et al., 2017)

Many studies, especially those conducted in the Arab world, still use the UK norms to interpret CPM to generate IQ scores (e.g. Bakhiet & Lynn, 2015; Bakhiet et al., 2017; Khaleefa & Lynn, 2008), often referred to as the “Greenwich IQ” (Bakhiet et al., 2018). However, this practice may be problematic due to the questionable appropriateness of using norms from another country to interpret test scores even for “culturally fair” tests (Daugherty et al., 2017; Norman et al., 2011). Additionally, some researchers have emphasized that one set of norms may not be applicable to the total population of a country especially for multicultural countries. Instead, there should be multiple normative options (Oliveri & von Davier, 2016; Solano-Flores & Li, 2013). The clinical implications relating to the misuse of and the unavailability of representative norms are extremely important, especially during a period of growth in migration and globalization in which we have reached the highest number of migrants in history (International Organization for Migration, 2018).

This study aims to examine the potential misclassification of intelligence in preadolescent Moroccan children when using non-representative norms to interpret the purportedly “culture free” CPM. Norms from UK, Spain, and Oman were selected due to contrasting cultural differences with Morocco. The first two countries belong to Western societies that differ substantially from Morocco in language, SES (World Bank, 2020), and quality of education (OECD, 2019; Walsh et al., 2010). Additionally, norms from these countries were selected for their clinical utility. CPM norms from the UK are the original and most widely used in previous research, whereas Spanish norms allow for practical application due to the large Moroccan immigrant population in this neighboring country (Migration Policy Institute, 2017). Oman is another Arab country sharing religion, language, values, and focus of education (El-Kogali& Krafft, 2019) with Morocco, although the economy is stronger. The inclusion of this Middle Eastern country provides a contrast with countries that are culturally similar (Buré-Reyes et al., 2013).

We hypothesize: 1) raw scores from the Moroccan children will differ from the normative data of the other three countries, and 2) IQ scores based on non-representative norms will result in an over-classification of Moroccan children in lower ranges versus the normal curve.

Method

Participants

One hundred forty-seven school children ages 7 (n = 51; 25 boys and 26 girls), 9 (n = 48; 23 boys and 25 girls), and 11 (n = 48; 22 boys and 26 girls) from two middle class schools in

Chefchaouen were included in this study. This northern city is representative of the culture and socio-economic conditions in Morocco which is based in tourism and trade (Fasfous et al., 2015). Children in each age group were similar in sociodemographics such as academic achievement, parental education, and parental employment status (Table 1). There was no evidence of cognitive, psychological, or physical impairment in the participants which is further supported by typical placement of Moroccan children with developmental problems in adapted schools. Demographic information was procured in interviews with parents and teachers.

A power analysis for a one-way analysis of variance was conducted using the following parameters: number of groups k = 3, Cohen's f = .4, significance level at .01, and a minimal power of 80%. The effect size corresponds to a large effect, according to Cohen (1988). An estimated sample size of at least n = 30 for each group was obtained. Thus, our sample provided adequate power for analysis.

Instrument

The Raven's Coloured Progressive Matrices (CPM) (Raven et al., 1998) is a nonverbal test of intelligence commonly used for the assessment of 5- to 11-year-old children, as well as elderly people. This task consists of 36 items of increasing difficulty, which include an incomplete pattern with six options below it, of which only one is correct. The total score is recorded and classified into ranges based upon percentiles: range I = intellectually superior (above the 95th percentile); range II = above average (between the 75th and 95th percentile); range III = on average (between the 25th and 75th percentile); range IV = below average (between the 5th and 25th percentile); range V = intellectually impaired (under the 5th percentile). Alternatively, IQ scores can be calculated from normative data. For this study, the norms from the UK (Raven et al., 1998), Spain (Raven et al., 2011) and Oman (Kazem et al., 2009) were used. While the reliability data for the British sample was unavailable, the CPM has shown adequate psychometric properties in Spanish children with a high Kuder–Richardson reliability index (K-21) equal to .92, a split-half reliability of .87, and a test-retest reliability of .71 (Gómez Fernández, 1982). For the Oman study, a Cronbach's alpha of .88, a split-half reliability of .78, and a test-retest reliability of .56 were reported (Kazem et al., 2009). By contrast, Cronbach's alphas of 0.88, 0.90, and 0.89 for the 7-, 9-, and 11-year-olds have been specifically presented. Considering these and other reliability analyses conducted on the CPM in other countries, an average of around 0.80 has been estimated (Cotton et al., 2005).

Procedure

The Delegation of Education of Chefchaouen selected two middle-class schools based on demographic and SES data. From each of the schools, two classes from grades 2 (7-years-old), 4 (9-years-old), and 6 (11-years-old) were selected. Applying stratified sampling method to class rosters, the sample was subdivided considering gender and academic achievement at three levels (low, medium, and high), and then the children were randomly drawn from each stratum. Testing was conducted by an Arabic-speaking neuropsychologist in a classroom at each of the schools during academic hours. All participants were assessed with the paper-pencil CPM that followed the instructions for the individual administration of this test (Raven et al., 1998). The total number of correct items was recorded. The Research Ethics Committee at the University of XXXX approved this study. Formal permission to conduct the study in the two schools was obtained from the Delegation of Education of the city and the schools' directors. The participation of the selected students was voluntary, and informed consent was obtained from parents.

Statistical Analysis

To compare CPM group differences for Moroccan children, chi squares were conducted to compare demographic data, and a one-way ANOVA was conducted with post hoc comparisons with a Bonferroni correction adjusted to $\alpha = .017$. Multiple one sample t-tests were performed to compare scores for Moroccan children with the other groups using their norms. As the means and standard deviations of the three age groups were not available from the UK norms, the following formula proposed by Luo et al. (2018) was used to estimate the means by each age group's percentiles: $M = [0.7+(0.39/n)] \times [(q_1 + q_3)/2] + [0.3 - (0.39/n)]m$, where q_1 is the first quartile (25th percentile), q_3 is the third quartile (75th percentile), and m is the median (50th percentile). For the SD estimations, the equation proposed by Wan et al. (2014) was applied: $SD = (q_3 - q_1)/\eta(n)$, where $\eta(n)$ is a varying value given in Wan et al. (2014). These equations have been proven to be the most accurate method of estimation for sample distributions close to normality (McGrath et al., 2020). To control for potential inaccuracies due to the need to use calculations compare the UK data, confidence intervals for the Moroccan scores were adjusted to 99%. Additionally, effect sizes were also reported using the Cohen's d (Cohen, 1988).

Finally, to determine misclassification, the norms from these three foreign countries were used to transform the Moroccan total CPM scores into ranges and IQ, especially focusing on the appearance of intelligence deficiencies.All statistical analyses were performed using the R statistical software (R Core Team, 2020), with the addition of the Basic Statistics and Data Analysis (BSDA) package (Arnholt & Evans, 2017) for the cross-cultural comparisons.

Results

Moroccan Data

No significant differences were found for parent's education and occupation status for the three age levels (see Table 1). The means and standard deviations of scores for the three age groups in Morocco are presented in Table 2. A one-way ANOVA reported significant differences between the three groups ($F = 21.15$, $p < .001$). Bonferroni post-hoc comparisons indicated significant differences between the 7- and 9-year-olds ($p < .001$) and the 7- and 11-year-olds ($p < .001$), in favor of the older groups (Table 2). However, no significant differences were found between children ages 9 and 11 ($p = .417$).

Table 1

Sociodemographic Variables of the Moroccan Sample

	Total (*N* = 147)	7-year-olds (*n* = 51)	9-year-olds (*n* = 48)	11-year-olds (*n* = 51)	Age differences
Gender (*n*)					
Boys	70	25	23	22	χ^2 = .103 *p* = .95
Girls	77	26	25	26	
Children's academic achievement[a] (*M, SD*)	6.93 (1.32)	6.84 (1.16)	7.07 (1.42)	6.88 (1.4)	*F* = .4 *p* = .667
Father's years of formal education(%)					
Less than 6 years	57.8%	58.8%	58.3%	56.2%	χ^2 = .67 *p* = .955
Between 6-12 years	25.9%	25.5%	22.9%	29.2%	
More than 12 years	16.3%	15.7%	18.8%	14.6%	
Mother's years of formal education (%)					
Less than 6 years	70.8%	68.6%	75%	68.7%	χ^2 = 5.07 *p* = .28
Between 6-12 years	22.4%	23.5%	14.6%	29.8%	
More than 12 years	6.8%	7.8%	10.4%	2.1%	
Father's employment status (%)					
Self-employment	29.3%	31.4%	25%	31.3%	χ^2 = 13.67 *p* = .322
Family Business	2.7%	3.9%	0%	4.2%	
Unskilled Manual Labor	14.3%	15.7%	8.3%	18.8%	
Manual Labor	2.7%	3.9%	0%	4.2%	
Domestic Labor	0.7%	0%	2.1%	0%	
Employee	35.4%	33.3%	39.6%	33.3%	
Unemployed/Housewife	15%	11.8%	25%	8.3%	
Mother's employment status (%)					
Self-employment	3.4%	2%	2.1%	6.3%	χ^2 = 10.15 *p* = .603
Family Business	1.4%	0%	2.1%	2.1%	
Unskilled Manual Labor	1.4%	0%	2.1%	2.1%	
Manual Labor	0.7%	2.7%	0%	0%	
Domestic Labor	6.1%	5.9%	4.2%	8.3%	
Employee	6.8%	3.9%	12.5%	4.2%	
Unemployed/Housewife	80.3%	86.3%	77.1%	77.1%	

[a] Grade point average (maximum of 10) during the academic year

Table 2

Raw Means, Standard Deviations, and a One-Way ANOVA between Age Groups in the Moroccan Sample

		CPM Total Score			ANOVA			
Age (years)	*n*	M	SD	95% CI	*F* ratio	*p*	Partial η^2	Posthoc
7	51	20.92	6.48	[19.1, 22.74]				
9	48	26.04	5.35	[24.49, 27.6]	21.15	< .001	.227	7 < 9* 7 < 11*
11	48	27.69	5.35	[26.5, 28.87]				

Note. CPM = Raven's Coloured Progressive Matrices; CI = confidence interval; ANOVA = analysis of variance.

* p < .001

Cross-Cultural Differences

Multiple one sample t-tests were conducted to compare the Moroccan data with normative data from the UK, Spain, and Oman (see Table 3). When comparing Moroccan and the UK children, significant differences were only found in the 11-year-old group with UK children demonstrating higher scores (t = -6.118, p <.001, d = 1.197). Significant differences were also found between Morocco and Spain, with higher scores for Spanish children in both 7-year-old (t = -2.965, p =003, d = .421) and 9-year-old groups (t = -2.785, p = .005, d = .417). No comparisons were made with the 11-year-old group, as there is no published data for this age group in Spain. Finally, when comparing the two Arab countries, differences were only found in the 9-year-old group with the Moroccan children scoring higher than the Omani group (t = 2.595, p = .01, d = .455).

Table 3

Cross-Cultural Comparisons between Morocco and UK, Spain, and Oman

	Morocco	UK[a]	Spain[b]	Oman[c]	Morocco vs. UK		Morocco vs. Spain		Morocco vs. Oman	
Age (years)	M (SD)	M (SD)	M (SD)	M (SD)	*t*	*d*	*t*	*d*	*t*	*d*
7	20.92 (6.48)	20 (4.57)	23.68 (6.53)	20.20 (6.50)	.849	.164	-2.965**	.421	.682	.118
9	26.04 (5.35)	27.64 (5.39)	28.32 (5.59)	23.25 (6.82)	-1.363	.298	-2.785**	.417	2.595**	.455
11[d]	27.69 (4.08)	32 (3.05)	-	26.39 (6.37)	-6.118***	1.197	-	-	1.34	.243

Note. UK = United Kingdom; d = Cohen's d effect size.

[a] Sample size for UK: n(7) = 55; n(9) = 37; n(11) = 55.

[b] Sample size for Spain: n(7) = 1373; n(9) = 1490.

[c] Sample size for Oman: n(7) = 137; n(9) = 154; n(11) = 177.

[d] No data published for the 11-year-olds in Spain.

*p < .05; **p < .01; ***p < .001.

Misclassifications

Norms from the UK, Spain, and Oman were used to transform the total CPM score of each subject into ranges through the procedure presented in manual (Raven, Raven, & Court, 1998). Frequencies were computed to pay explicit attention to the percentage of subjects falling into the ranges IV ("below average", under the 25th percentile) and V ("intellectual impaired", under the 5th percentile) (Figure 1). IQ was also calculated using the norms of the three foreign countries for each age group (Figure 2).

Applying the Norms of UK

When the original norms from the UK were applied to the 7-year-old children, 23.53% were classified as "below average" (range IV), and 15.69% of them were classified as "intellectually impaired" (range V). In the 9-year-old group, 29.17% fell "below average" (range IV), and 6.25% presented as “intellectually impaired” (range V). Finally, 62.5% of the 11-year-olds were classified as “below average” (range IV) and 10.4% as “intellectually impaired” (range V). Using the same UK norms, the IQ mean for the 7-year-old group was 103, 96 for the 9-year-old group, and 79 for the 11-year-olds.

Applying the Norms of Spain

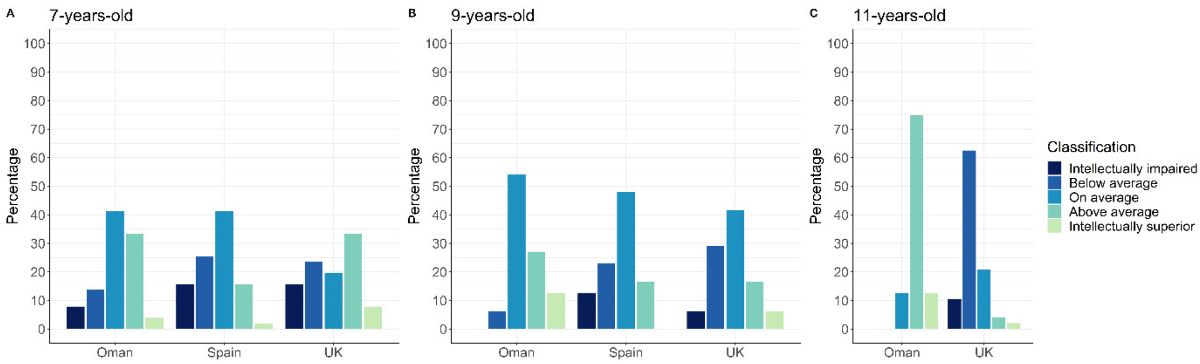
When the norms from Spain were applied to the Moroccan 7-year-old group, 25.49% of the children fell “below average” (range IV) and 15.68% were classified as “intellectually impaired” (range V). For the 9-year-old group, 22.92% were classified as “below average” (range IV) and 12.5% were “intellectually impaired” (range V).When calculating IQs, both 7- and 9-year-old Moroccan children earned an IQ score of 94.

Applying the Norms of Oman

Classification using Omani norms resulted in 13.73% of the 7-year-old sample falling in the “below average” (range IV) and 7.84% in the “intellectually impaired” (range V). The lowest classification for the 9-year-olds was “below average” (range IV) with 6.25% of children falling in this range. No 11-year-old Moroccan child scored lower than “average” (range IV). Intelligence scores were 102, 106, and 103 points for the 7-, 9- and 11-year-old Moroccan samples, respectively.

Figure 1

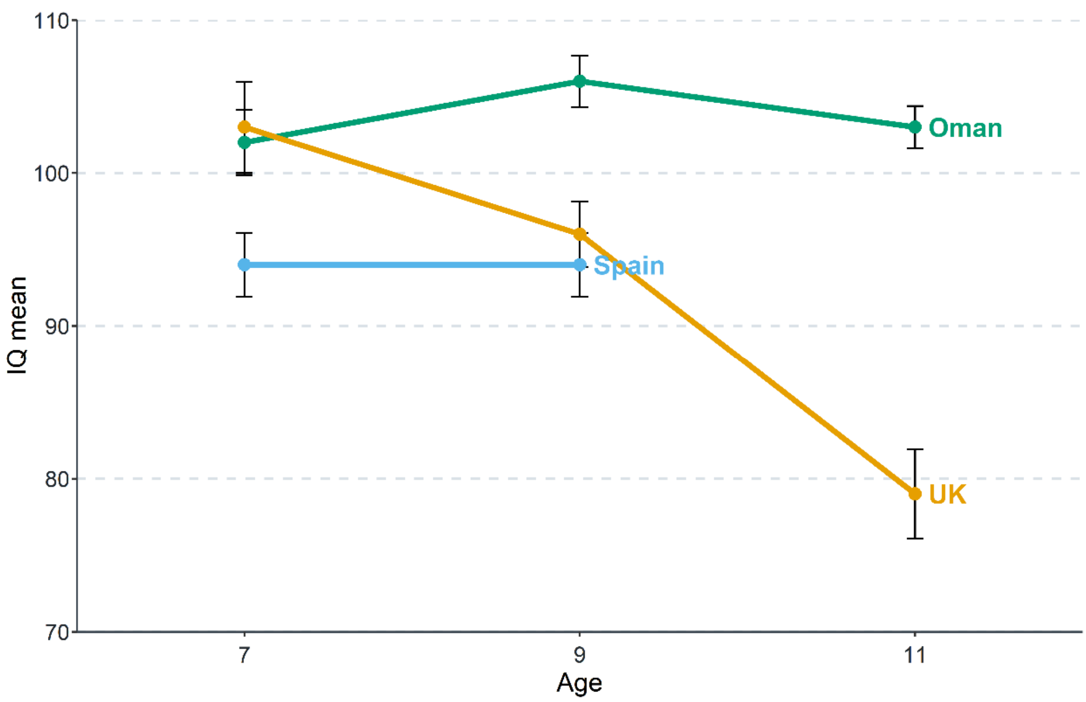
Misclassifications when Transforming the Moroccan Raw Scores into Ranges Applying the Norms from UK, Spain, and Oman



Note. No data published for the 11-year-olds in Spain. UK = United Kingdom.

Figure 2

Moroccan Raw Scores Transformed into IQ Applying the Norms from the UK, Spain, and Oman



Note. No data published for the 11-year-olds in Spain. UK = United Kingdom.

Discussion

Both of our hypotheses concerning potential biases in the “cultural fair” CPM when using non-representative norms were supported. The raw scores of Moroccan children were significantly lower than children from the UK (11-year-olds) and Spain (7- and 9-year-olds). This pattern of findings is consistent with performances on other international tests such as the Programme for International Student Assessment (PISA; OECD, 2019), and the Progress in International Reading Literacy Study (PIRLS; Martin et al., 2017), where Moroccan children perform lower than those from the UK and Spain.

Two related cultural factors that could be impacting CPM scores are economics and quality of education (Manly et al., 2002). According to the World Bank (2020), the UK and Spain are high-income countries, whereas Morocco’s economy falls within the low to middle category. A country’s economy has been correlated with IQ scores on western tests (Meisenberg,2012). Economics can significantly impact quality of education, as countries with the strongest economies tend to have the best educational infrastructure such as highly rated schools, educational expenditures per student, and broadband use, thus would likely provide the best learning opportunities (McPhillips, 2017; OECD, 2016; World Bank, 2020). In addition, educational curriculum could be another contributing factor as pedagogies in Morocco are generally focused on rote-learning (El-Kogali& Krafft, 2019; Wagner, 1993), while other Western countries, such as the UK or Spain, place more emphasis on active play-based and exploratory learning (Walsh et al., 2010).

An interesting finding when using UK norms, is the lack of significant differences in raw scores for the 7- and 9-year-old cohorts, but significantly lower scores for the 11-year-olds. A possible explanation for this differential pattern is the long-term impact of educational quality on cognitive development. Studies have indicated that years of education are a significant contributor to intelligence development (Cahan & Cohen, 1989; Ritchie & Tucker-Drob, 2018) and quality of education can have long-term impact on cognition (Manly et al., 2002). Thus, it would be plausible that longer periods of exposure to stronger education would result in better scores on intellectual testing over time, as seen in our UK and Moroccan data. The stagnation of Moroccan scores between 9- and 11-year-olds would be supportive of this interpretation. More research isneeded to examine the long-term effects of educational quality.

In contrast to UK and Spanish children, Moroccan children significantly outperformed the Omani 9-year-old group. This result is counterintuitive, given the cultural similarities and Oman's stronger economy andeducational ranking (World Bank, 2020; United Nations Development Programme, 2020). However,studies have demonstrated that cognitive differences exist between countries that are culturally similar, for example in language, religion, and educational style (El-Kogali& Krafft, 2019; Porcaro, 2011). Our findings would indicate that there are unknown cultural factors aside from educational and economic factors accounting for performances on western tests. One key may be geographic proximity, as geographicdistance between countries has been found to be highly correlated with IQ (Gelade, 2008).

Our second hypothesis, IQ scores based on non-representative norms will result in an over-classification of Moroccan children in lower ranges versus the normal curve, was partially supported. We found a large percentage of children falling "below average" (range IV) and classified as "intellectually impaired" (range V) when using the norms from both the UK and Spain. These findings have important implications for the assessment of both Moroccans living in their country and abroad. This test, together with the already limited neuropsychological tools that have been properly adapted and validated in Morocco (Fasfous et al., 2017), can potentially lead to errors of measurement and test interpretation when using non-representative norms. Additionally, considering the proximity of Morocco to Spain and the large number of migrants reaching 707,000 by 2017 (Migration Policy

Institute, 2017), there is also a high probability for misclassification of Moroccan children emigrating to Spain who are assessed for clinical or educational reasons.

Clinicians should also be cautious when using the original normative CPM data (UK) to calculate the IQ of the Moroccan children. While the Moroccan children obtained higher raw scores at older ages, the opposite effect was shown when these scores were transformed into a Greenwich IQ with mean IQs decreasing from “below average” at age 7 to “intellectually impaired” at age 11 (almost 2 SDs under the mean). Given similar findings, Bakhiet et al. (2018) concluded that IQ decreases with age in the Arab world. Our findings would suggest that this conclusion was erroneous based upon use of non-representative norms for data interpretation and support Greenfield’s (1997) determination that psychological tests may not cross cultures.

Notwithstanding, this study has some limitations. First, is the small sample size and limited number of age groups, particularly for the Spanish group. A broader age range with larger samples would be important to ensure confidence in our pattern of findings. Second, our cross-sectional design did not control for possible cohort effects which could account for our findings. Third, weaknesses in the test-retest reliability of the Omani norms would limit the validity of our findings. Finally, the demographic data for the study groups were insufficient, which would limit our understanding of group equivalency and group characteristics on test performance.

To increase the accuracy of intellectual assessment when assessing children from different cultures, future studies should develop tests for specific cultures that follow the International Test Commission guidelines for adapting tests (International Test Commission, 2017) and gather norms for sub-cultures or minority groups within the same population (Oliveri & von Davier, 2016). In addition, examining the impact of contextual factors, for example, SES, educational differences, religion, values, behaviors, or economics, on performance on western intelligence tests can both help to elucidate influences on intellectual development and interventions to improve cognition.

Acknowledgements

Special thanks to Julia C. Daugherty and CharitiniPitsiakou for their early contribution to the improvement of themanuscript.

Funding

This work was supported by the Centro de Iniciativas de Cooperación al Desarrollo (CICODE) [project PI18X].

Conflict of Interest

None of the authors has any conflicts of interest to disclose.

References

Agranovich, A. V., Panter, A. T., Puente, A. E., &Touradji, P. (2011). The culture of time in neuropsychological assessment: Exploring the effects of culture-specific time attitudes on timed test performance in Russian and American samples. *Journal of the International Neuropsychological Society, 17*(4), 692-701. https://doi.org/10.1017/S1355617711000592

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review, 15*(4), 185. http://doi.org/10.1007/s11065-005-9180-y

Ardila, A., Bertolucci, P. H., Braga, L. W., Castro-Caldas, A., Judd, T., Kosmidis, M. H., ... & Rosselli, M. (2010). Illiteracy: the neuropsychology of cognition without reading. *Archives of clinical neuropsychology, 25*(8), 689-712. https://doi.org/10.1093/arclin/acq079

Arnholt, A. T., & Evans, B. (2017). *BSDA: Basic Statistics and Data Analysis*. R package version 1.2.0. https://CRAN.R-project.org/package=BSDA

Bakhiet, S. F. A., Abdelrasheed, N. S. G., Cheng, H., Lynn, R., Essa, Y. A. S., &Blahmar, T. A. M. (2017). Sex Differences on the Coloured Progressive Matrices in Sudan. *Mankind Quarterly, 57*(4), 581.

Bakhiet, S. F. A., Dutton, E., Ashaer, K. Y. A., Essa, Y. A. S., Blahmar, T. A. M., Hakami, S. M., & Madison, G. (2018). Understanding the Simber Effect: Why is the age-dependent increase in children's cognitive ability in Arab countries than in Britain?.*Personality and Individual Differences, 122*, 38-42. https://doi.org/10.1016/j.paid.2017.10.002

Bakhiet, S. F. A., & Lynn, R. (2015). Regional differences in intelligence in Sudan. *Intelligence, 50*, 150-152. https://doi.org/10.1016/j.intell.2015.03.007

Bandeira, D. R., Alves, I. C. B., Giacomel, A. E., &Lorenzatto, L. (2004). Matrizesprogressivas coloridas de Raven-escala especial: normas para Porto Alegre, RS. *Psicologia em estudo, 9*(3), 479-486. https://doi.org/10.1590/S1413-73722004000300016

Buré-Reyes, A., Hidalgo-Ruzzante, N., Vilar-López, R., Gontier, J., Sánchez, L., Pérez-García, M., & Puente, A. E. (2013). Neuropsychological test performance of Spanish speakers: Is performance different across different Spanish-speaking subgroups?.*Journal of Clinical and Experimental Neuropsychology, 35*(4), 404-412. https://doi.org/10.1080/13803395.2013.778232

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child development, 60*(5), 1239-1249. https://doi.org/10.2307/1130797

Cattell, R. B. (1940). A culture-free intelligence test. I. *Journal of Educational Psychology, 31*(3), 161. https://doi.org/10.1037/h0059043

Cattell, R. B. (1973). *Culture-fair intelligence test*. Institute for personality and ability testing.

Cattell, R. B., Feingold, S. N., &Sarason, S. B. (1941). A culture-free intelligence test: II. Evaluation of cultural influence on test performance. *Journal of Educational Psychology, 32*(2), 81. https://doi.org/10.1037/h0058456

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R., &Crewther, S. G. (2005). A normative and reliability study for the Raven's Coloured Progressive Matrices for primary school aged children from Victoria, Australia. *Personality and individual differences, 39*(3), 647-659. https://doi.org/10.1016/j.paid.2005.02.015

Daugherty, J. C., Puente, A. E., Fasfous, A. F., Hidalgo-Ruzzante, N., & Pérez-Garcia, M. (2017). Diagnostic mistakes of culturally diverse individuals when using North American neuropsychological tests. *Applied Neuropsychology: Adult, 24*(1), 16-22. https://doi.org/10.1080/23279095.2015.1036992

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13-21. https://doi.org/10.1016/j.intell.2006.02.001

El-Kogali, S. E. T., & Krafft, C. (Eds.). (2019). *Expectations and Aspirations: A New Framework for Education in the Middle East and North Africa*. The World Bank. http://doi.org/10.1596/978-1-4648-1234-7

Fasfous, A. F., Al-Joudi, H. F., Puente, A. E., & Pérez-García, M. (2017). Neuropsychological measures in the Arab world: A systematic review. *Neuropsychology review, 27*(2), 158-173. http://doi.org/10.1007/s11065-017-9347-3

Fasfous, A. F., Puente, A. E., Pérez-Marfil, M. N., Cruz-Quintana, F., Peralta-Ramirez, I., & Pérez-García, M. (2013). Is the color trails culture free? *Archives of clinical neuropsychology, 28*(7), 743-749. https://doi.org/10.1093/arclin/act062

Fasfous, A. F., Peralta-Ramirez, M. I., Pérez-Marfil, M. N., Cruz-Quintana, F., Catena-Martinez, A., & Pérez-García, M. (2015). Reliability and validity of the Arabic version of the computerized Battery for Neuropsychological Evaluation of Children (BENCI). *Child Neuropsychology, 21*(2), 210-224.

Flynn, J. R. (2012). *Are we getting smarter?: Rising IQ in the twenty-first century*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139235679

Fujii, D. (2017). *Conducting a culturally informed neuropsychological evaluation*. American Psychological Association. https://doi.org/10.1037/15958-000

Gelade, G. A. (2008). IQ, cultural values, and the technological achievement of nations. *Intelligence, 36*(6), 711-718.https://doi.org/10.1016/j.intell.2008.04.003

Georgas, J. (2003). Cross-cultural psychology, intelligence, and cognitive processes. In J. Georgas, L. G. Weiss, F. J. van de Vijver, & D. H. Saklofske (Eds.), *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (pp. 23-37). San Diego, CA: AcademicPress.

Gómez Fernández, D. (1982). Propiedades psicométricas del" PMC-Raven" a partir de respuestas emitidas por escolares gallegos de EGB. *Revista de Psicología General y Aplicada*.

Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American psychologist, 52*(10), 1115. https://doi.org/10.1037/0003-066X.52.10.1115

Grigorenko, E. L., Meier, E., Lipka, J., Mohatt, G., Yanez, E., & Sternberg, R. J. (2004). Academic and practical intelligence: A case study of the Yup'ik in Alaska. *Learning and Individual Differences, 14*(4), 183-207.https://doi.org/10.1016/j.lindif.2004.02.002

International Organization for Migration. (2018). *Global Migration Indicators*. https://publications.iom.int/system/files/pdf/global_migration_indicators_2018.pdf

International Test Commission. (2017). *The ITC guidelines for translating and adapting tests*. (2nd ed.) https://www.InTestCom.org/

Kazem, A. M., Alzubiadi, A. S., Alkharusi, H. A., Yousif, Y. H., Alsarmi, A. M., Al-Bulushi, S., … &Alshammary, B. M. (2009). A normative study of the Raven Coloured Progressive Matrices Test for Omani children aged 5-11 Years. *Jurnal Pendidikan Malaysia| Malaysian Journal of Education, 34*(1), 37-51.

Kellogg, C. E., & Morton, N. W. (1999). *Beta III*. San Antonio: Psychological Corporation.

Khaleefa, O., & Lynn, R. (2008). A study of intelligence in the United Arab Emirates. *Mankind quarterly, 49*(1), 58-64.

Linstrom, A., Raven, J., & Raven, J. (2008). The coloured progressive matrices in South Africa. In J. Raven, & J. Raven (Eds.), *Uses and abuses of intelligence. Studies advancing Spearman and Raven's quest for non-arbitrary metrics* (pp. 363-370). Royal Fireworks Press.

Luo, D., Wan, X., Liu, J., & Tong, T. (2018). Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. *Statistical methods in medical research, 27*(6), 1785-1805. https://doi.org/10.1177%2F0962280216669183

Lynn, R., Abdalla, S., & Al-Shahomee, A. A. (2008). Norms for the progressive matrices for Libya and Tunisia. *Mankind Quarterly, 49*(3), 71-77.

Martin, M. O., Mullis, I. V., & Hooper, M. (2017). *Methods and Procedures in PIRLS 2016*. International Association for the Evaluation of Educational Achievement.

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society: JINS, 8*(3), 341.https://doi.org/10.1017/S1355617702813157

McPhillips, D. (2017). Best countries for education: *From primary school to university, a look at how countries invest in the world's future leaders*. U.S. News.https://www.usnews.com/news/best-countries

Meisenberg, G. (2012). National IQ and economic outcomes. *Personality and Individual Differences, 53*(2), 103-107.https://doi.org/10.1016/j.paid.2011.06.022

Melikyan, Z. A., Puente, A. E., &Agranovich, A. V. (2020). Cross-Cultural Comparison of Rural Healthy Adults: Russian and American Groups. *Archives of Clinical Neuropsychology*. https://doi.org/10.1093/arclin/acz071

McGrath, S., Zhao, X., Steele, R., Thombs, B. D., & Benedetti, A. (2020). Estimating the sample mean and standard deviation from commonly reported quantiles in meta-analysis. *Statistical Methods in Medical Research.* https://doi.org/10.1177/0962280219889080

Migration Policy Institute. (2017). *Immigrant and Emigrant Populations by Country of Origin and Destination*. https://www.migrationpolicy.org/programs/data-hub/charts/immigrant-and-emigrant-populations-country-origin-and-destination

Norman, M. A., Moore, D. J., Taylor, M., Franklin Jr, D., Cysique, L., Ake, C., ... &Hnrc Group. (2011). Demographically corrected norms for African Americans and Caucasians on the hopkins verbal learning test–revised, brief visuospatial memory test–revised, stroop color and word test, and wisconsin card sorting test 64-card version. *Journal of clinical and experimental neuropsychology, 33*(7), 793-804. https://doi.org/10.1080/13803395.2011.559157

Oliveri, M. E., & von Davier, A. A. (2016). Psychometrics in support of a valid assessment of linguistic minorities: Implications for the test and sampling designs. *International Journal of Testing, 16*(3), 220-239. http://doi.org/10.1080/15305058.2015.1069743

Organisation for Economic Cooperation and Development (OECD). (2016). *Broadband statistics update*. https://www.oecd.org/digital/broadband-statistics-update.htm

Organisation for Economic Cooperation and Development (OECD). (2019). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing.

Porcaro, D. (2011). Reviewing the literature of computer-supported collaborative learning (CSCL) to determine its usefulness in Omani education development. *International Journal of Education and Development using ICT, 7*(3).

Puente, A.E., & Perez-Garcia, M. (2000). Neuropsychological assessment of ethnic minorities: Clinical issues. In F.A. Paniagua (Ed.), *Handbook of multicultural mental health* (pp. 419–435). San Diego, CA: Academic Press.

R Core Team (2020). *R: A language and environment for statistical computing*. http://www.R-project.org/

Raven, J. C. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press.

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive psychology, 41*(1), 1-48. https://doi.org/10.1006/cogp.1999.0735

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: The Coloured Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

Raven, J. C., Court, J. H., Raven, J. (Eds.). (2011). *Raven: Matrices progresivas. Escalas Color (CPM). General (SPM). Superior (APM)*. Madrid: Pearson.

Rindermann, H. (2008). Relevance of education and intelligence at the national level for the economic welfare of people. *Intelligence, 36*(2), 127-142. https://doi.org/10.1016/j.intell.2007.02.002

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological science, 29*(8), 1358-1369. https://doi.org/10.1177/0956797618774253

Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and cognition, 52*(3), 326-333. https://doi.org/10.1016/S0278-2626(03)00170-2

Shuttleworth-Edwards, A. B., Kemp, R. D., Rust, A. L., Muirhead, J. G., Hartman, N. P., & Radloff, S. E. (2004). Cross-cultural effects on IQ test performance: A review and preliminary normative indications on WAIS-III test performance. *Journal of clinical and experimental neuropsychology, 26*(7), 903-920. https://doi.org/10.1080/13803390490510824

Solano-Flores, G., & Li, M. (2013). Generalizability theory and the fair and valid assessment of linguistic minorities. *Educational Research and Evaluation, 19*(2-3), 245-263. https://doi.org/10.1080/13803611.2013.767632

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. Cambridge: Cambridge University Press.

Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist, 59*, 325–338. https://doi.org/10.1037/0003-066X.59.5.325

Sternberg, R. J. (2007). Intelligence and Culture. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 547-568). New York, NY: Guilford Press.

Sternberg, R. J., Nokes, C., Geissler, P. W., Prince, R., Okatcha, F., Bundy, D. A., & Grigorenko, E. L. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence, 29*(5), 401-418.https://doi.org/10.1016/S0160-2896(01)00065-4

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*(5), 401-426. https://doi.org/10.1016/j.intell.2006.09.004

UnitedNationsDevelopment Programme. (2020). *Human Development Report 2020*. http://hdr.undp.org/en/2020-report

von Stumm, S., & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence, 48*, 30-36. https://doi.org/10.1016/j.intell.2014.10.002

Wagner, D. A. (1993). *Literacy, culture and development: Becoming literate in Morocco*. Cambridge University Press.

Walsh, G., Sproule, L., McGuinness, C., Trew, K. and Ingram, G. (2010) *Developmentally Appropriate Practice and Play-Based Pedagogy in Early Year's Education: A Literature Review of Research and Practice*. School of Psychology, Queen's University Belfast Stranmillis University College, Belfast.

Wan, X., Wang, W., Liu, J., & Tong, T. (2014). Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC medical research methodology, 14*(1), 135. http://doi.org/10.1186/1471-2288-14-135

World Bank. (2020). *World Development Indicators*. https://databank.worldbank.org/

Ziada, K. E., Metwaly, H. A. M., Bakhiet, S. F. A., Cheng, H., & Lynn, R. (2017). A new study of intelligence in Egypt. *Mankind Quarterly, 58*(2), 307-311.