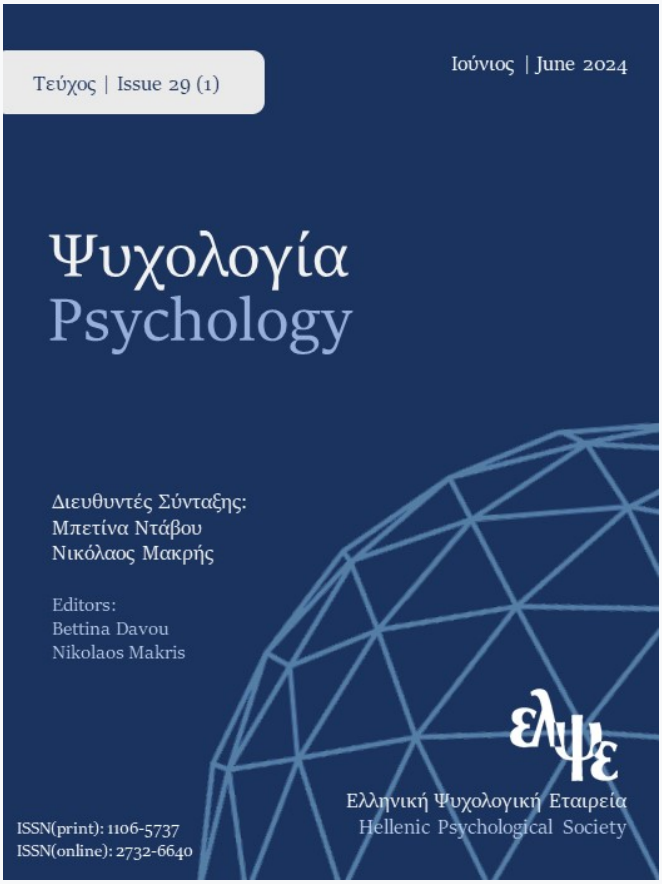
Psychology: the Journal of the Hellenic Psychological Society

Vol 29, No 1 (2024)

June 2024



Assessment of adult cognitive abilities in Greece: A differential item functioning study of the General Ability Measure for Adults (GAMA)

Gary Rempe, Achilles N. Bardos, Zoe Bablekou, Panagiotis Simos, Maria Doropoulou

doi: 10.12681/psy_hps.34184

To cite this article:

Rempe, G., Bardos, A. N., Bablekou, Z., Simos, P., & Doropoulou, M. (2024). Assessment of adult cognitive abilities in Greece: A differential item functioning study of the General Ability Measure for Adults (GAMA). *Psychology: The Journal of the Hellenic Psychological Society*, *29*(1), 231–240. https://doi.org/10.12681/psy_hps.34184

ΨΥΧΟΛΟΠΑ | PSYCHOLOGY, 29(1), 231-240

Ελληνική Ψυχολογική Εταιρεία
Hellenic Psychological Society

ΕΜΠΕΙΡΙΚΗ ΕΡΓΑΣΙΑ | RESEARCH PAPER

Assessment of adult cognitive abilities in Greece: A differential item functioning study of the General Ability Measure for Adults (GAMA)

Gary P. REMPE[1], Achilles N. BARDOS[2], Zoe BABLEKOU[3], Panagiotis SIMOS[4], Maria DOROPOULOU[5]

[1] Colorado Center for Assessment & Counseling
[2] EduMetrisis, Limited Liability Company
[3] Child Psychology & Education Lab, School of Early Childhood Education, Faculty of Education, Aristotle University of Thessaloniki
[4] School of Medicine, University of Crete
[5] Department of Psychology, Panteion University of Social and Political Sciences

KEYWORDS

Intelligence assessment
Intercultural assessment
Differential element/question function

CORRESPONDENCE

Achilles N. Bardos, Ph.D.
EduMetrisis, LLC
P.O. BOX 336404
Greeley, CO 80633 -USA
abardos@comcast.net

ABSTRACT

The purpose of this study was to evaluate the generalizability and possible adaptation for use of a non-verbal measure of intelligence developed in the United States, the General Ability Measure for Adults (GAMA; Naglieri & Bardos, 1997) in Greece. As a validity evidence the study examined the differential item function of its questions in order to explore potential item bias utilizing the disproportionate group probabilities of participants' correctly endorsing test items. The analysis was performed using a logistic regression procedure with samples from the United States (n = 2,369) and Greece (n = 1,273). The findings indicate a small (<1%) number of items that work differentially between these two cultural groups. Implications for the development and weighting of cross-cultural intelligence assessment tests using non-verbal measures are discussed.

Introduction

Numerous assessment instruments have been developed in Greece (Stalikas et al., 2012) some of which constitute an original work of authorship but the majority are translations and adoptions of tools developed in other countries, mostly in the United States. Moreover, few of them investigate bias and report related findings, a necessary piece of validity evidence. As regards the assessment of cognitive abilities for adults in Greece, there is only one standardized tool currently available for this purpose, the Wechsler Adult Intelligence Scale - 4th Edition (WAIS-IV GR) (Wechsler, 2014). Standardization for the Greek WAIS-IV was carried out by the Motibo Publishing company under the direction of Professor Stogiannidou (2014) with the assistance of NCS Pearson Inc. The Greek WAIS-IV was standardized using a nationally representative sample of 895 participants (16:0-90:11 years), stratified into 13 age categories, which matched closely the gender distribution in the Greek general population. Participants were also stratified into three education categories: (a) Elementary (0-9 years of education), (b) secondary (10-12 years), and (c) higher education (13+ years). The need for additional measures of cognitive ability with established psychometric properties is apparent and further necessitated as the Wechsler Scales are heavily influenced by verbal skills and, despite the efforts by researchers to adapt to local linguistic demands, this remains a challenge and weakness in clinical evaluation(s).

Psychology: The Journal of the Hellenic Psychological Society
https://ejournals.epublishing.ekt.gr/index.php/psychology
https://doi.org/10.12681/psy_hps.34184

Measures with reduced or very limited linguistic requirements have been proposed (Green et al., 2016) as being more sensitive/equitable and more appropriate in clinical evaluations of cognitive ability. As the ethnic and linguistic composition of the Greek population has been changing during the last decade, an assessment tool utilizing non-verbal stimuli like the GAMA (Naglieri & Bardos, 1997) appears to be a viable option to consider in clinical evaluations. A number of studies in Greece have utilized the GAMA as a measure of cognitive ability in clinical populations (Simos et al., 2014; Spyridaki et al., 2014), and human resource settings (Lemonaki et al., 2021). However, an extensive examination of the instrument's item bias through a differential item functioning analysis (DIF) has not been conducted, and this is therefore the current study's goal. We hypothesized that, given the non-verbal nature of the GAMA, a very small number of items might be found where the populations of United States and Greece might be different.

Item bias

Notable differences in group performance on high publicity assessments during the mid 1900s gave rise to public outcry regarding fairness in testing across various groups of people. As a reaction, stringent methods for identifying bias and selecting new items in testing were outlined in the Golden Rule Settlement (1984), the result of a five year court battle between the Educational Testing Services (ETS) and the Golden Rule Insurance Company in the USA. However, the psychometric community expressed concerns about the methods employed for identifying and addressing these items, as they would likely lead to invalidate the tests themselves (Bond, 1987). Following these concerns, new methodologies were developed, many by the ETS, to investigate item equity across different groups of people. Currently, inquiries into potential item bias are an essential step in any assessment validation process (International Testing Commission, 2001). These investigations, also known as differential item functioning (DIF) analyses, attempt to uncover disproportionate probabilities of correctly endorsing test items that may exist between any sampling groups. While the initial categories of interest in DIF analyses were primarily ethnicity-based, this work has been extended to include gender and country membership, among other groups.

Recent decades have seen many developments in DIF methodologies (Walker, 2011). While there is not a clear consensus in the literature as to which method is the most effective, practitioners seem to favor the Mantel-Haenszel procedure (MH; Holland & Thayer, 1988) and Logistic Regression (LR; Swaminathan & Rogers, 1993). Though both methods have been shown to effectively identify uniform DIF, LR has the additional capacity to test for non-uniform DIF. Recent work in the area of DIF analysis has highlighted the importance of examining non – uniform DIF (Ong et al., 2015).

Non-uniform DIF can be thought of as an interaction between ability level and group membership that contributes to different probabilities for correct item responses. A critical aspect of DIF analysis is the matching or controlling of ability levels when assessing group probabilities. In item response theory (IRT) methodology, this can be visualized through an examination of item characteristic curves (ICC; see Figures 1 to 4). It is critical to explore non-uniform DIF, given that this intersection of the two group curves can obscure the statistical test for uniform DIF. Li and Stout (1996) recommend that it is only when the crossing point approaches the middle of the ability level that the uniform test will be compromised and refer to this as true 'crossing DIF', with all ancillary intersections termed directional non-uniform DIF.

In DIF analysis it is common to examine several different groups within a sample. These groups have historically included groups based on ethnicity and gender, but recent cross-cultural work has highlighted the importance of examining culturally influenced DIF, particularly in assessments that are used internationally and without the generation of new normative data. Unfortunately, the latter occurrence is relatively commonplace and may contribute to misleading test results, if differences in test validity that may occur when tests are adapted across cultures are not ruled out (Roivainen, 2013).

Of interest in this study was the General Ability Measure for Adults, a brief non-verbal measure of cognitive functioning (GAMA; Naglieri & Bardos, 1997) and its viability as a measure of cognitive ability for adults in Greece. This study uses GAMA's standardization data from the United States and a large sample from Greece to explore cross-national DIF. As the GAMA was designed specifically to reduce confounds such as access to formal education and linguistic background, researchers hypothesize findings will indicate low amounts of DIF when comparing the Greek sample to the United States.

The findings reported in this paper stem from a pilot study in which the MH procedure was used. Preliminary results indicated low amounts of uniform DIF, however, Breslow Day Test of Homogeneity statistics suggested the possibility of a large amount of non-uniform DIF in many items (Penfield, 2003). The purpose of this study is to corroborate the initial MH findings, and to further explore potential non-uniform relationships using the LR procedure. Moreover, this paper aims to contribute to research on cross-national DIF, and to explore the international utility of a non-verbal intelligence test in Greece.

Method

Participants

The sample for this study included 2,369 participants from the United States aged 18-96 and 1,273 Greek participants aged 18-94. Both samples were stratified in terms of education, geographical region, and clinical population. GAMA administration in Greece was conducted in the participants' native language. The GAMA includes very minimum verbal instructions when introducing the test's sample items, which were translated and back translated from English to Greek. The United States sample was part of the test's standardization while the Greek sample was gathered over a period of five years (2013-2018) and it is currently undergoing further psychometric analysis for the establishment of Greek norms (See Table 1 for descriptive information about each sample).

Table 1. *GAMA Sample Data by Country (USA, Greece)*

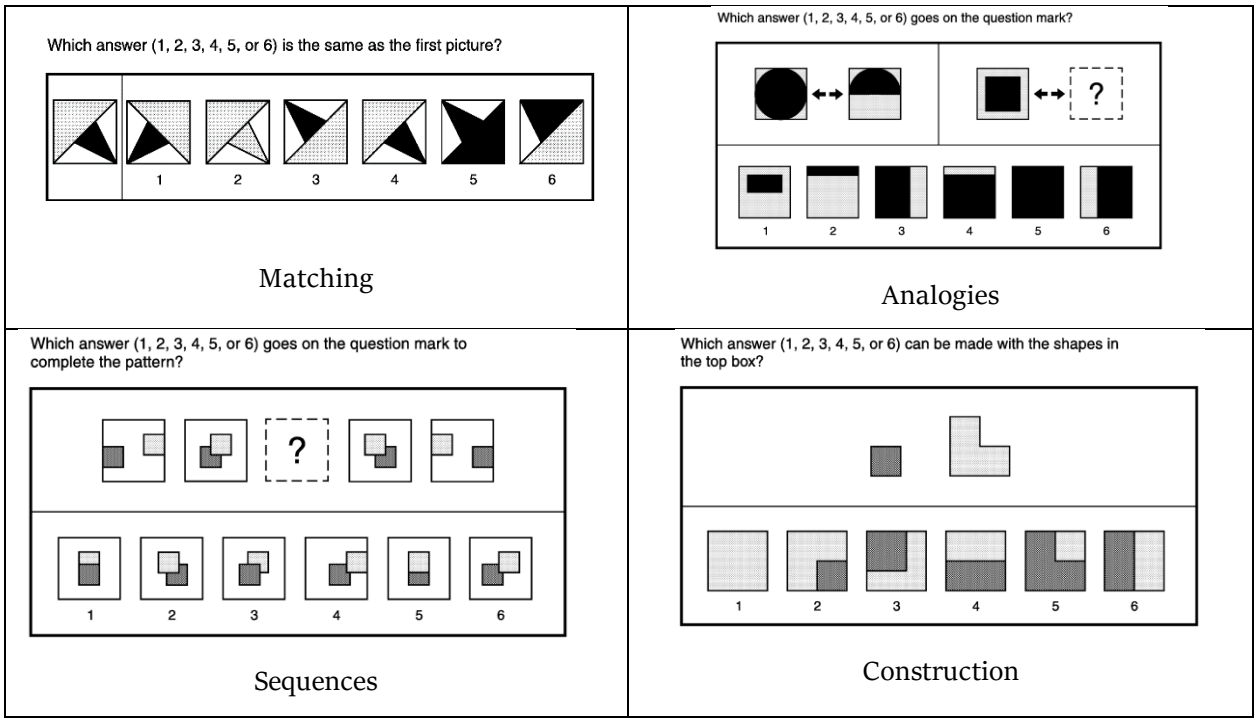
		Gender		**Age**	**Ability**
Country	*n*	Male	Female	Mean (*SD*)	Mean (*SD*)
United States	2360	1079	1281	46.44(*20.84*)	30.16(*11.31*)
Greece	1373	607	766	37.72(*16.53*)	35.72(*11.87*)

Tool and Procedure

The GAMA is composed of 64 non-verbal items that can be administered within a 25-minute time period individually or in a group setting using the paper and pencil version or through a web-based platform. Each item is represented by colorful geometric designs that require no verbal response. The Cronbach's alpha overall internal consistency is excellent for both the United States (a = .94) and the Greek sample (a = .94). The GAMA consists of four subtests with a mean of 10 and SD=3, each contributing equally to the overall GAMA Total score which is presented with a mean of 100 and SD =15. The four subtests are not to be considered measures of different abilities but rather different means of measuring general ability. Their different content represents an effort to maintain the interest of the examinee in the assessment process. Figure 1 presents some sample items from each subtest. Please note that all test items incorporate colors in their actual format. On the Matching subtest, examinees select stimuli that are alike in color, shape, and configuration and then are required to select which of six options presented is identical in shape, color, and configuration. Analogies items require the examinee to identify the relationship between two abstract figures and then find a different pair of figures that

completes another relationship. On the Sequence subtest, a group of geometric designs form a logical sequence of patterns with a missing part that the examinee is required to recognize and complete within the sequence. Finally, on the Construction subtest, the examinee is required to mentally synthesize two or three designs to create one that is offered as one of the options.

Figure 1. *Sample items of the GAMA subtests*



Data analysis

The LR procedure for testing DIF compares the fit of a series of nested models; a model indicating no DIF, uniform DIF, or non-uniform DIF. Prior to analysis, all item responses are dichotomously coded with 0 being the focal group (Greece) and 1 being the reference group (US). The grouping variable, country in this case, is also dichotomously coded with 1 being the US or reference group, and (0=zero) being Greece or the focal group. We treated the missing data as a random event so no statitistical manipulation was performed. Further missing data are systematically omitted in the ability level variable and the parameter estimations. The ability level is a simple raw score summation. The logistic regression models employed in this study can be described as follows:

$$\text{Model 1: } \text{Ln}\left(\frac{P\text{mi}}{1-P\text{mi}}\right) = \beta 0 + \beta 1(\text{total score})$$

$$\text{Model 2: } \text{Ln}\left(\frac{P\text{mi}}{1-P\text{mi}}\right) = \beta 0 + \beta 1(\text{total score}) + \beta 2(\text{country})$$

$$\text{Model 3: } \text{Ln}\left(\frac{P\text{mi}}{1-P\text{mi}}\right) = \beta 0 + \beta 1(\text{total score}) + \beta 2(\text{country}) + \beta 3(\text{total score} * \text{country})$$

$\text{Ln}\left(\frac{P\text{mi}}{1-P\text{mi}}\right)$ denotes the logistic function in which *Ln* is the natural logarithm of the probability of person m correctly endorsing item i (Pmi), over the probability of person m incorrectly endorsing item i (1 – Pmi). Model 1 is the null model in which $\beta 0$ represents the constant and total score or $\beta 1$ represents ability level, also referred to as the conditioning variable. By including the grouping variable, Model 2 tests for uniform DIF (a significant $\beta 2$ value). Model 3, by including the interaction term $\beta 3$, tests for non–uniform DIF. Non – uniform DIF indicates

that as ability is varied, there are differing probability values for the reference and focal groups correctly endorsing an item (see Figures 2-5).

The exponential of the grouping parameter $\beta 2$ ($\hat{a}_{\mathrm{LR}}$), has been shown to be equivalent to the common odds estimate given in the MH procedure ($\hat{a}_{\mathrm{MH}}$; Monahan et al., 2007). $\beta 2$ can also be thought of as the slope of the item characteristic curve or the discrimination parameter. The statistic $\hat{a}_{\mathrm{LR}}$ provides an odds ratio estimate of the group of interest correctly endorsing an item for every one unit increase in the predictor. A value greater than 1 indicates DIF disadvantaging the focal group, and a value below 1, DIF disadvantaging the reference group. $\hat{a}_{\mathrm{LR}}$ can be further transformed to the delta scale used in the MH procedure to obtain an effect size (Δ_{LR}; Monahan, et al.., 2007). This transformation yields a more interpretable value with 0 representing equal group probabilities, by taking its natural log and then multiplying it by -2.35; Δ_{LR} = -2.35(ln($\hat{a}_{\mathrm{LR}}$). After the transformation, a positive value indicates DIF disadvantaging the reference group, and a negative disadvantaging the focal group. To add meaning to these numbers, Dorans and Holland (1993), working with the Educational Testing Service (ETS), developed a three-level classification system of DIF type based on the MH procedure. This classification system has been adapted to the LR procedure and the significance test is derived by comparing the likelihood ratio values between the different nested models, which fall on a chi- squared distribution with 1 degree of freedom:

Type A: Negligible - X^2 p = 0, $a \leq .001$ OR $|\Delta_{\mathrm{LR}}| < 1$

Type B: Moderate - X^2 $p \neq 0$, $a \leq .001$ OR $|\Delta_{\mathrm{LR}}| > 1 < 1.5$

Type C: High - X^2 $p \neq 0$, $a \leq .001$ AND $|\Delta_{\mathrm{LR}}| > 1.5$

Meaningful effect sizes have not yet been developed for items in which the full model is significant, indicating non-uniform DIF (i.e., an interaction between ability level and country). To explore these items, a series of logistic graphs were generated, using saved predicted probabilities and ability levels, for visual inspection of the relationships. Figures 2-4 show examples of the four possible outcomes from the LR procedure using results from this analysis. The Y-axis was obtained by saving predicted probabilities derived from the LR procedure in SPSS, version 19. The different item characteristic curves were generated with the $\beta 2$ parameter from the second logistic model.

Due to the multiple hypotheses being tested in this analysis, a conservative p-value was used for the model comparison using the Bonferroni type adjustment. A 99% confidence level was determined to sufficiently and reasonably control error rates: $.05/66 \sim p < .001$. Following the flagging of items containing high amounts of DIF, the analysis was conducted again with these items excluded, a process known as purification of the matching criterion (Zumbo, 1999).

Results

There were 13 items displaying statistically significant uniform DIF at the 99% confidence level. Using the ETS effect size system, 10 of these items were type B and the remaining three fell in the type C category. Every item containing moderate to high DIF showed negative Δ_{LR} values, indicating a lower probability of correctly endorsing the item for the focal group (Greece).

Breaking the analysis up by subtest, Matching showed the least amount of DIF with only 1 type B item out of 11 total. Analogies was close behind with 2 type B items out of 17 total. Sequences showed two type B items and 1 type C item out of 20 total items. The subtest with the highest number of items containing DIF was Construction, with 5 type B and two type C items out of 18 total items.

Concerning non-uniform DIF, there were 10 items for which the third model was significant at the 99% confidence level. Four of these items also showed significant uniform DIF, 2 at type B and 2 at type C. By subtest, Matching displayed no non-uniform DIF, Analogies 1 item, Sequences 4 items, and Construction 5 items.

Following visual inspection of predictive probability graphs for these items, 9 of the 10 items show the Greek sample crossing the United States at a point well above theta (see Figure 4) indicating directional non – uniform DIF (Li & Stout, 1996). This left only one item showing significant crossing DIF (item displayed in Figure 5), item 34, belonging to the Sequences subscale (See Table 2 for the results of the DIF analysis).

Figures 2-5. *Examples of the four possible outcomes from the LR procedure using results from this analysis*

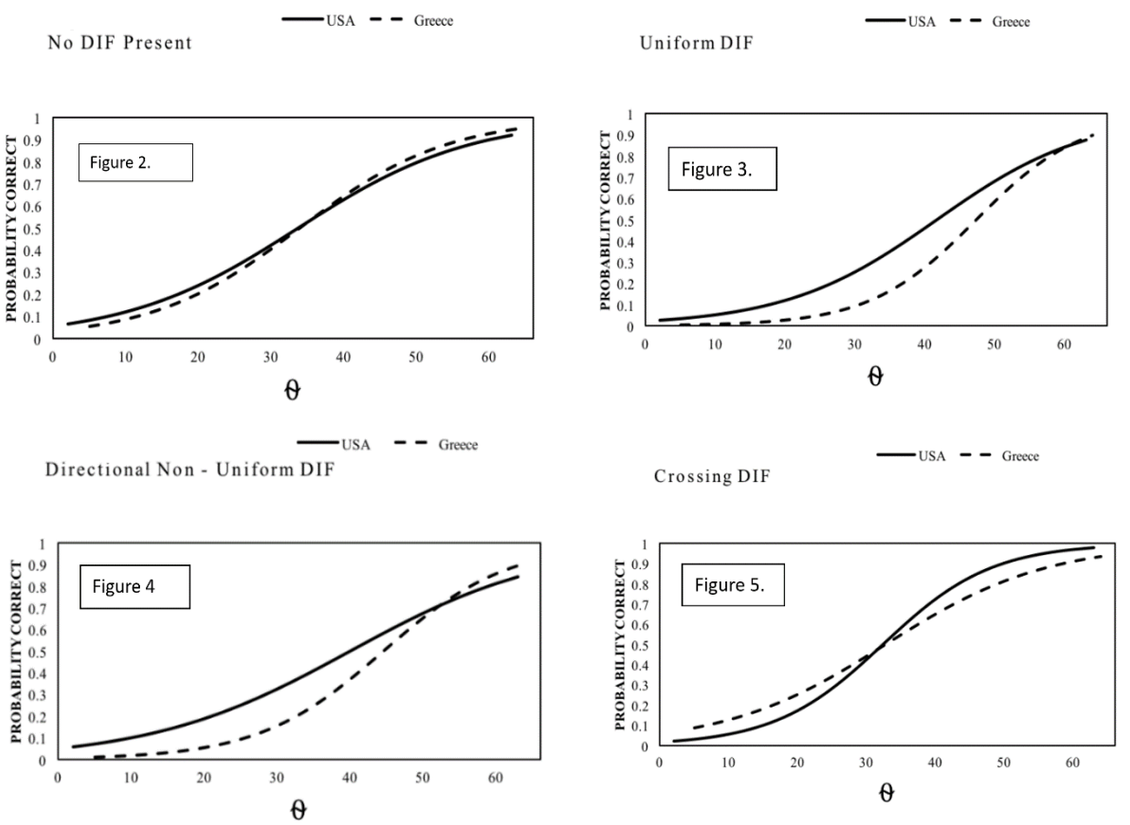


Table 2. *Results of DIF Analysis by Country (USA, Greece)*

Subtest	Item Number	$\hat{a}_{LR}$	Δ_{LR}	ETS Rating	Disadvantaged Country
Matching	38	1.463	-0.895	B	Greece
Analogies	37	1.624	-1.140	B	Greece
	56	1.855	-1.452	B	Greece
Sequences	21	1.423	-0.830	B	Greece
	34*	1.140	-0.308	A	Greece
	39	1.581	-1.076	B	Greece
	61	2.397	-2.054	C	Greece
Construction	30	1.366	-0.733	B	Greece
	40	1.595	-1.097	B	Greece
	43	1.618	-1.130	B	Greece
	48	1.710	-1.262	B	Greece
	52	1.717	-1.269	B	Greece
	58	2.396	-2.054	C	Greece
	59	2.249	-1.904	C	Greece

**Note.* * = item is recommended for deletion

Discussion

This study used data from two large and stratified samples of individuals from Greece and the United States on a non-verbal IQ test to explore the possibility of culturally induced item-wise bias. The purpose of using LR was to corroborate earlier MH findings from a pilot study, and to explore non-uniform relationships. LR analysis findings indicate the largest amount of DIF in the Construction subscale.

Construction subtest items asked examinees to mentally combine shapes to form new geometric designs. This task involves spatial reasoning, working memory, and abstract reasoning. On every item that was flagged with moderate to high DIF, Greece had a lower probability of correctly endorsing an item. This indicates the subtest may be measuring a secondary dimension in the Greek sample, however, it may also be that this is inconsequential DIF, as discussed by Angoff (1993). It is important to highlight that only two of the items actually fall into the high DIF category, a relatively small number.

Given the differences in sample collection date, it is important to consider these findings in light of the Flynn Effect (Flynn, 1984; Trahan et al., 2014). According to Flynn's observations, we would expect to see a standard three-point increase per decade. Indeed, the mean Greek score is five points, a third of a standard deviation, higher than the United States sample. If this were impacting DIF statistics, it seems that it would be the Greek sample that showed the higher probabilities, not the United States. However, it may also be true that the effect is masking DIF by bringing the item probabilities closer than they otherwise would be. For this to occur, there would need to be extreme amounts of DIF in the test. This is theoretically unjustified, but a limitation of the study regardless.

As the GAMA was designed specifically to reduce confounds such as access to formal education and linguistic background, researchers hypothesized findings would indicate low amounts of DIF. The retention of this hypothesis may have implications for other international assessments that are non-verbal in nature. The standardization process is both effortful and expensive, the results of this study may warrant further questioning into the position that all assessments undergo a full norming process when being adapted across cultures especially when receptive and expressive language skills are clearly not a possible confounding variable and theoretically and rationally generalize well to the new culture. Perhaps other psychometric techniques of building equivalent test forms might be a good alternative.

Overall, the MH findings were corroborated using LR and the GAMA shows low amounts of uniform DIF in the Greek sample. Utilizing the directional non-uniform DIF framework outlined by Li and Stout (1996), only one item shows relevant crossing DIF. Adding this item to the type C items brings the total high DIF items up to 4 out of 66. It appears that the GAMA, possibly due to its non-verbal design, may be a reasonable candidate for use in Greek populations without the need to create separate norms. However given that the GAMA was administered to a large Greek sample (1,273 Greek participants aged 18-94), stratified in terms of education, geographical region, Greek independent norms can be created while further studies should explore the performance of clinical populations along with studies that will examine the presence of DIF on the GAMA using alternate cultural groups in Greece, particularly those in which the differences from western culture are more substantial.

References

Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3-23). Lawrence Erlbaum.

Bond, L. (1987). The golden rule settlement: A minority perspective. *Educational Measurement: Issues and Practice, 6*(2). 10-20. https://doi.org/10.1111/j.1745-3992.1987.tb00406

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Lawrence Erlbaum.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932–1978. *Psychological Bulletin, 95*(1). 29–51. https://doi.org/10.1037/0033-2909.95.1.29

Green, A., Bardos, A. N., & Doropoulou, M. (2016). The General Ability Measure for Adults-second edition. In R.S. McCallum. *Handbook of Nonverbal Assessment*. pp 207–216. Kluwer Academic.

Golden Rule Insurance Company v. Washburn, 419-76 Illinois Cir. Ct., 7th Ind. Cir. Ct. (1984). Consent decree.

Holland, P. W., & Thayer, D. (1988). Differential item performance and the Mantel Haenszel procedure. In H. Weiner & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Lawrence Erlbaum.

International Test Commission. (2001). International guidelines for test use, *International Journal of Testing, 1*(2), 93-114. https://doi.org/10.1207/S15327574IJT0102_1

Lemonaki, R., Xanthopoulou, D., Bardos, A.N., Karademas, E., & Simos, P. (2021). Burnout and job performance: a two-wave study on the mediating role of employee cognitive functioning. *European Journal of Work and Organizational Psychology, 30*(3), 1-13. https://doi.org/10.1080/1359432X.2021.1892818

Li, H. H., & Stout., W. (1996). A new procedure for detection of crossing DIF. *Psychometrica, 61*(4), 647-677. https://eric.ed.gov/?id=EJ540314

Monahan, P., McHorney, C., Stump T., & Perkins A. (2007). Odds ratio, delta, ETS classification, and standardization measures of DIF magnitude for binary logistic regression, *Journal of Educational and Behavioral Statistics*, *32*(1), 92 – 102. https://doi.org/10.3102/1076998606298035

Naglieri, J. & Bardos, A. (1997). *General Ability measure for Adults*. National Computer Systems.

Ong, Y., M., Lamprianou, I., & Williams J. (2015). Exploring crossing differential item functioning by gender in mathematics assessment. *International Journal of Testing, 15*(4), 337-355. https://doi.org/10.1080/15305058.2015.1075639

Penfield, R. D. (2003). Applying the Breslow-day test of trend in odds ratio heterogeneity to the analysis of nonuniform DIF. *Alberta Journal of Educational Research, 49*(3), 231-243. https://psycnet.apa.org/record/2003-10912-002

Roivainen, E. (2013). Are cross-national differences in IQ profiles stable? A comparison of Finnish and U.S. WAIS norms. International Journal of Testing, 13(2), 140-151. https://doi.org/10.1080/15305058.2012.657315

Simos, P., Papadopoulou, S., Vakis, A., Tsagaraki, P., Papastefanakis, E. & Bardos, A. N. (2014). Anterograde verbal episodic memory deficits after traumatic brain injury: The effects of trauma severity and presence of executive deficits. *Hellenic Journal of Psychology*, *11*(1), 1-25.

Spyridaki, E., Simos, P., Avgoustinaki, P., Demirtzaki, E., Venihaki, M., Bardos, A.N. & Margiotis, A. (2014). The association between obesity and fluid intelligence impairment is mediated by chronic low grade inflammation. *British Journal of Nutrition. 112,* 1724-1734 https://doi.org/10.1017/S0007114514002207

Stalikas, A., Triliva, S., & Roussi, P. (2012). Ta psychometrica ergaleia stin Ellada [Psychometric tools in Greece]. Pedio.

Swaminathan, H., & Rodgers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27,* 361–270. https://doi.org/10.1111/j.1745-3984.1990.tb00754.x

Trahan, L., Stuebing, K. K., Hiscock, M. K., & Fletcher, J. M. (2014). The Flynn Effect: A Meta-analysis. Psychological Bulletin, 140(5), 1332–1360. https://doi.org/10.1037/a0037173

Walker, C. M. (2011). What's the DIF? why differential item functioni analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment, 29*(4), 364-376. https://doi.org/10.1177/0734282911406666

Wechsler, D. (2014). *WAIS IV GR Klimaka Noimosinis gia enilikes WECHSLER –Tetarti ekdosi: Egxeiridio xorigisis kai aksiologisis* NCS Pearson, (Greek Edition) [*WAIS-IV GR-Wechsler Adult Intelligence Scale - 4th Edition Administration and Scoring Manual*]. MOTIBO.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense.

ΕΜΠΕΙΡΙΚΗ ΕΡΓΑΣΙΑ | RESEARCH PAPER

Αξιολόγηση των γνωστικών ικανοτήτων ενηλίκων στην Ελλάδα: Μελέτη της διαφορικής λειτουργίας των ερωτήσεων στην Κλίμακα Γενικής Νοημοσύνης Ενηλίκων (GAMA)

Gary P. REMPE[1], Αχιλλέας Ν. ΜΠΑΡΔΟΣ[2], Ζωή ΜΠΑΜΠΛΕΚΟΥ[3], Παναγιώτης ΣΙΜΟΣ[4], Μαρία ΝΤΟΡΟΠΟΥΛΟΥ[5]

[1] Colorado Center for Assessment & Counseling
[2] EduMetrisis, Limited Liability Company
[3] Εργαστήριο Ψυχολογίας & Εκπαίδευσης του Παιδιού-*ΨΥΧΗ*, Τμήμα Προσχολικής Αγωγής και Εκπαίδευσης, Παιδαγωγική Σχολή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
[4] Ιατρική Σχολή, Πανεπιστήμιο Κρήτης
[5] Τμήμα Ψυχολογίας, Πάντειο Πανεπιστήμιο Κοινωνικών και Πολιτικών Επιστημών

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Αξιολόγηση νοημοσύνης
Διαπολιτισμική αξιολόγηση
Διαφορική λειτουργία στοιχείων/ερωτήσεων

ΣΤΟΙΧΕΙΑ ΕΠΙΚΟΙΝΩΝΙΑΣ

Achilles N. Bardos, Ph.D.
EduMetrisis, LLC
P.O. BOX 336404
Greeley, CO 80633 -USA
abardos@comcast.net

ΠΕΡΙΛΗΨΗ

Ο σκοπός αυτής της μελέτης ήταν να αξιολογήσει την πιθανή χρήση μιας μη-λεκτικής κλίμακας γενικής νοημοσύνης για ενήλικες (GAMA; Naglieri & Bardos, 1997) στην Ελλάδα, ερευνώντας τη διαφορική λειτουργία των ερωτήσεων. Η ανάλυση πραγματοποιήθηκε χρησιμοποιώντας τη διαδικασία λογιστικής παλινδρόμησης με συμμετέχοντες/ουσες από τις Ηνωμένες Πολιτείες (n =2,369) και την Ελλάδα (n = 1,273). Η μεθοδολογία της στατιστικής ανάλυσης χρησιμοποιεί τις δυσανάλογες ομαδικές πιθανότητες των συμμετεχόντων να απαντήσουν σωστά στις ερωτήσεις της κλίμακας GAMA. Τα ευρήματα από την ανάλυση δείχνουν έναν μικρό αριθμό στοιχείων (<1%) που λειτουργούν διαφορικά μεταξύ αυτών των δύο πολιτισμικών ομάδων. Συζητιούνται οι επιπτώσεις σχετικά με την ανάπτυξη και στάθμιση διαπολιτισμικών δοκιμασιών αξιολόγησης της νοημοσύνης χρησιμοποιώντας μη-λεκτικές μετρήσεις.

https://doi.org/10.12681/psy_hps.34184

Ψυχολογία: Το Περιοδικό της Ελληνικής Ψυχολογικής Εταιρείας
https://ejournals.epublishing.ekt.gr/index.php/psychology