Research Article

OMICS International

Matrix Analogies Test - Short Form: An Evaluation of its Measurement Properties for a Sample of Greek School-aged Children using Rasch Analysis

Georgia Spiliotopoulou[*]

Department of Clinical Sciences, College of Health and Life Sciences, Brunel University, Mary Seacole Building, Uxbridge, Middlesex, UB8 3PH, United Kingdom

[*]**Corresponding author:** Georgia Spiliotopoulou, Department of Clinical Sciences, College of Health and Life Sciences, Brunel University, Mary Seacole Building, Uxbridge, Middlesex, UB8 3PH, United Kingdom, Tel: (0044)18952 68827; Fax: (0044)18952 69853; E-mail: georgia.spiliotopoulou@brunel.ac.uk

Abstract

Objective: Matrix analogies test–short form (MAT-SF) is a screening test of non-verbal reasoning originally validated with US students. The purpose of this study was to investigate its construct validity and reliability with Greek school-aged children using Rasch analysis.

Methods: Data were collected from 106 typically developing children aged 7-11 years old and were Rasch-analysed using RUMM2030.

Results: On removal of seven items, the Rasch-scaled 27-item MAT-SF demonstrated successful overall (item-trait interaction χ^2 (27)=52.282; p=0.0024, bonferonni adjustment α=0.0018) and individual fit to the model. There was no differential item functioning for gender or age or disordered response thresholds and unidimensionality was demonstrated. The person-separation reliability (PSI=.832) indicated the tool's ability to discriminate between three different groups. The tool was well targeted to the group of typically developing children. Significant differences between 7-9 and 9-11 years groups (F(1,101)=13.53; p<0.0001) and between children with higher and lower reading ability (F(1,101)=43.82; p<0.0001) further supported validity.

Conclusion: It is concluded that the revised MAT-SF demonstrates a justifiable research and screening tool of non-verbal reasoning to be used with professionals working with school-aged children in Greece.

Introduction

Fluid intelligence (Gf) is conceptualised as a non-verbal, abstract, independent of previous knowledge, and relatively culture free aspect of general intelligence [1-3]. Gf has been found to be the best predictor of performance in situations that involve human intelligence including performance at school, university and in cognitively demanding situations [4]. In children it is a predictor of a wide range of cognitive abilities with low Gf predicting academic difficulties [5-7].

Taub et al. [7] demonstrated statistically significant direct effects of Gf on mathematic achievement of participants aged 5-19 years. Previous studies present similar findings [8,9]. More specifically Gf is considered to account for some of the prominent problem-solving constructs, accurate numerical calculation and strategies implicated in mathematics performance [7,10,11]. With respect to the prediction of reading performance there are no definite findings. Alloway et al. [5] showed that non-verbal intelligence was significantly but weakly associated with teachers' assessments on reading ability of 194 children aged 4-5 years. However, Gathercole et al. [12] could not identify non-verbal intelligence as a mediator of working memory in predicting reading achievement. They attributed these findings to the low level of non-verbal intelligence in the participants. Moreover, Floyd et al. [13] did not find important influences of the abilities associated with Gf on reading decoding skills of participants aged 5 to 39 years old. They suggested that when a full range of cognitive abilities including general

intelligence is considered, Gf is overshadowed by more important influences.

Research in children has also shown that Gf is linked to working memory, which in turn impacts on children's academic achievement [14-17]. The strength of the correlations between Gf and working memory varies depending on the measurement method used in these studies. Heitz et al. [18] explained this relationship through the controlled attention mechanism.

Gf has been empirically defined as the latent trait extracted from a variety reasoning-dominated tests [19]. Testing involves presenting humans with abstract problems that are not likely to have seen before suggesting that successful performance cannot be attributed to previous learning [20]. Matrix analogies test – short form (MAT-SF) [21] is one of the screening tools for non-verbal reasoning in children. Although standardised in the USA, it has been found to be easily adapted for use with children of different socio-cultural backgrounds, as it requires minimal verbal instructions, demands no verbal expressive skills and it is less culturally loaded [21]. This is very important in the case of Greece, where the field of intellectual assessment is underdeveloped and there is a great need for standardising such foreign instruments. This would allow professionals working in school settings or conducting research for psychoeducational purposes to screen for non-verbal reasoning and predict the capacity to perform successfully in a wide range of situations and academic achievement [20,22].

Petrogiannis et al. [22] made a first attempt to look at the criterion and construct validity of MAT-SF with 731 children of various ages from five cities in Greece. They examined the relationships between MAT-SF and an index of educational performance referring to the average of all academic courses taken. They also looked for gender differences and for performance differences between Greek and US population. Correlations with age and developmental changes were also investigated. Their results were similar to the original standardisation sample with moderate but statistically significant correlations (r=0.30, p<0.01) between MAT-SF and academic achievement and lack of gender differences. Correlations with age were significant but lower in strength and the performance of the Greek sample indicated differences for two (8 and 9 years old) out of the 12 age groups in comparison to the US sample. They concluded that MAT-SF can be used as a screening tool using the US norms until Greek norms become available. Since then, no other studies have been conducted to look further at the validity and reliability of MAT-SF with Greek school aged children.

More recent developments in the measurement theory suggest the use of Rasch analysis to assess existing tools that are intended to be summated into an overall score allowing for improvements to their structure. Rasch analysis tests the reliability and validity of a tool, as this has currently been redefined into a unitary concept, named as construct validity. This assesses the ability of the items included in an instrument, collectively and individually, to reflect a single cohesive theoretical concept that they aim to measure and looks at the trustworthiness of score meaning and its interpretation [23,24]. The scales are tested against the expectations of the Rasch model, which is considered to be a template that operationalises the formal axioms which underpin measurement [25]. MAT-SF has not been Rasch analysed neither with the US nor with a Greek sample. Considering some of the differences identified in the Greek study (especially with 8 to 9 years old children), the lack of reliability testing and the lack of factor analysis to check for a unidimensional tool in its use with Greek children, there is a need to Rasch analyse MAT-SF to provide further evidence for its validation in Greece.

The purpose of this feasibility study is to extend the research of Petrogiannis et al. [22] in order to see whether MAT-SF meets the formal requirements of measurement as defined by the Rasch model to be used for screening of non-verbal reasoning with Greek school-aged children. The aim is to look at the construct validity and reliability of the tool with a group of typically developing children 7-11 years old in Greece considering item reduction if necessary. A test of reading comprehension [26] was also applied to see whether MAT-SF was able to discriminate between higher and lower reading ability.

Materials and Methods

Participants

A sample of 106 typically developing children aged 7-11 years were drawn from 10 urban mainstream primary schools to represent the schools in Attiki, Greece as a whole in terms of socio-cultural status. Children with major neurological or sensory deficits, or not native speakers, as was indicated by the class teacher, were excluded. There were no further selection constraints as it was considered that children attending mainstream schools would be typically developing in general.

The mean chronological age of the sample was 8 years and 9 months (M=105.34 months; SD=13.207). For the analysis they were grouped into two age groups. Fifty-eight children were included in the 7-9 years group (24 girls, 34 boys) and 48 in the 9-11 years group (32 girls, 16 boys). Group comparisons indicated that children's mean chronological age was very similar to their mean non-verbal intelligence age (M=103.60 months; SD=29.426), indicating that they can be considered as typically developing in general.

Instruments

Matrix analogies test - short form (MAT-SF): MAT-SF [21] consists of 35 colourful designs (the first one is used for practice) in a matrix format with missing elements in each item. It includes four kinds of items (i.e. pattern completion, reasoning by analogy, serial reasoning and spatial visualisation), which appear in a varying order. MAT-SF can be administered as a quick screening test of non-verbal reasoning to identify children who may have learning problems on the basis of an ability/achievement discrepancy or intellectually gifted children [21]. The scores derived from MAT-SF should be used to enable decisions as to whether additional testing is needed [27].

The test was standardised with 4468 US students aged 5-17 years. Its construct validity was checked with factor analysis for each age group. Concurrent validity was demonstrated by significant correlations (p>. 01) with the Multilevel academic survey test for mathematics and reading (MAST) [28], and with non-verbal ability as measured with WISC-R (r=0.68, p<0.001) with students with hearing difficulties [29]. Its internal consistency (Chronbach's α=0.83), as well as its test-retest reliability (0.51<r<0.91) and the standard error of measurement were considered as satisfactory for a screening test. Other studies have looked at its concurrent validity (r=0.73) with the Stanford-Binet intelligence scale: Fourth edition [30], and the Draw-A-Person: A quantitative scoring system in learning disabled (r=0.42) and non-handicapped students (r=0.50) [31]. Significant correlations were also shown between MAT-SF and academic achievement [30-33].

Screening test for reading ability [26]: This test was used to measure reading comprehension using a multiple choice format with 42 uncompleted sentences. The child had to choose one out of four given words to complete each sentence. The test was validated for use with school-aged Greek children and it had very good internal consistency (Chronbach's α=0.94) and split-half reliability (Guttman's α=0.93). Its standard error of measurement was 2.42. Criterion validity was assessed by correlating performance on the test with the subjective assessment of school teachers on reading comprehension (0.26<r<1.00, p<0.001). Developmental changes were also demonstrated.

Procedure

The screening tests were administered within the school premises in a separate and quiet room by the same researcher. Each child participated in two different group-based sessions, one for the assessment of non-verbal reasoning (25 min) and one for the assessment of reading comprehension (40 min). Written consent from the parents and the Head of each school, as well as assent consent from the children were obtained. Ethical approval was granted by the Hellenic Ministry of Education.

Data analysis

Data were Rasch evaluated using the RUMM2030 software to test for the internal construct validity of MAT-SF and check if the

construct can be improved. Rasch analysis shows what should be an expected pattern of responses to items if interval scale measurement is to be achieved [34]. The model assumes that the probability of a respondent to the scale affirming an item is a logistic function of the difference between the person's ability and the difficulty of the item on a linear scale [34]. Therefore, the Rasch model builds a hypothetical unidimensional line on which items and persons are located according to their difficulty and ability measures [35].

The unrestricted model was used because the likelihood-ratio test was statistically significant (p<0.001). The data were tested for (a) appropriate category ordering. One would expect that as person ability increases, the likelihood of obtaining a higher score increases as well in a logical order. Disordered thresholds occur when participants have difficulty in consistently discriminating between response options [36]; (b) the fit of the data to the model. Three overall fit statistics were considered. Two were item-person interaction statistics transformed to approximate a z-score, representing a standardised normal distribution. If the items and persons fit the model, one would expect approximately M=0, SD=1. The third statistic was an item-trait interaction reported as chi-square reflecting the property of invariance across the trait. A non-significant chi-square (p >0.05) indicates no substantial deviation from the model, whereas a significant chi-square suggests that the hierarchical ordering of the items varies across the trait. Additionally, individual item- and person- fit statistics were checked, where fit residual values outside the ± 2.5 range, or p values smaller than the bonferonni adjustment level indicate misfit to the model [36]; (c) unidimensionality, which is required for a valid summed raw score. This was tested by using item fit statistics and principal component analysis (PCA) of the residuals and by looking for the presence of local dependency of the residuals [37]; (d) differential item functioning (DIF) to show whether the different subgroups (e.g. age and gender) in the sample respond in a different manner to an individual item, despite equal levels of the underlying characteristic being measured [34,36]; (e) the scale's ability to target the population being assessed and for its reliability using the person-separation index (PSI). This indicates how well the items of the tool separate or spread out the subjects in the sample. A PSI of 0.70 represents the ability to distinguish two distinct strata of person ability [38].

Validity was also checked by looking at the ability of the tool to differentiate between different age groups and between children of higher and lower reading ability, what is defined in classic measurement theory as criterion validity.

Results

Initial overall fit

Of the 35 items, item I001 was excluded from the analysis as it was just for practice. Item I002 and one child (girl, 7-9 age groups) were removed as extreme scores. The initial fit of the data to the Rasch model showed a significant item-trait interaction χ^2 (3)=98.023, p<0.001 suggesting misfit between the data and the model. The mean (SD) fit residual values were -0.122 (1.431) for items and -0.037 (0.813) for persons showing some misfit to the model for the items. The mean (SD) location of the persons was -0.162 (1.165) suggesting that in general the response group was of a slight lower ability level than the difficulty level of MAT-SF. The PSI of reliability was .858, which means that the test had the power to discriminate among three groups of participants, which is considered as excellent power.

Estimates of person and item measures

Individual person-fit statistics showed that one girl (7-9 age groups) had fit residual of 3.568. In subsequent analyses, on removal of persons outside the accepted range (>2.5), two more children (girls, 7-9 age group) were identified (2.857 and 2.673) and removed. This resulted in 102 children remaining in the analysis and in improved means of person's fit residual (-0.037 to -0.006), item's fit residual (-0.122 to -0.117) and person's location (-0.162 to -0.099). The item-trait interaction remained significant nonetheless. The individual item-fit statistics showed that items I023, I035 and I028 were outside the accepted range (± 2.5) with the first two items having p values smaller than the bonferroni adjusted α<0.0015. On gradual removal of these items, there was further improvement in the means (SD) of item's fit residuals and person's location to -0.086 (1.026) and -0.035 (1.026) respectively. The item-trait interaction was also improved, but remained significant (χ^2 (30)=72.331, p=0.000023, bonferroni adjustment α=0.001667).

Unidimensionality testing

Local dependency was checked by looking at the residuals correlations by adding 0.03 at the average inter-item correlation of -0.03. The items above the value of 0.27 were I008 with I010 and I016 with I019. This indicates that the response on one of these items determines the response on its pair item. As a result items I019 and I008 were gradually deleted as this provided a better solution than deleting their pair-items. Following this, borderline local dependency was also identified between items I015 and I016 leading to removal of I016 as a better alternative.

PCA was conducted for the remaining 27 items. The first factor of the PCA is the primary contributor to the variance of the data, when the Rasch factor is not taken into account. The items that were loading more strongly upon this factor were I026, I005, I010, I025, I008, I019, I006, I013, I033, I011, I020, I029 (positively), and I003, I009, I017, I028, I002, I024, I030, I014, I023, I032, I012, I016 (negatively). These two sets of different person estimates were compared using paired t-test analyses. For the 102 tests performed, 6 (5.94%) were significant (p<.05). Since the proportion of significant tests was over 5%, a binomial test was conducted to provide a defined range of what is acceptable amount of deviating results given the sample size. The binomial test showed that the lower 95% confidence interval fell at .017, which is lower than .05 and indicates acceptable unidimensionality.

Following the above item removals, the item-trait interaction χ^2 (27)=52.282 was no more significant (p=0.002455, exceeding the bonferonni adjustment α=0.001852). The means (SD) for the item-fit, person-fit residuals and person location were -0.080 (0.853), -0.068 (0.444) and -0.136 (1.224), respectively.

Category ordering

Looking at figure 1, the category ordering demonstrated ordered thresholds without need for any rescoring from the first to all subsequent Rasch analyses.
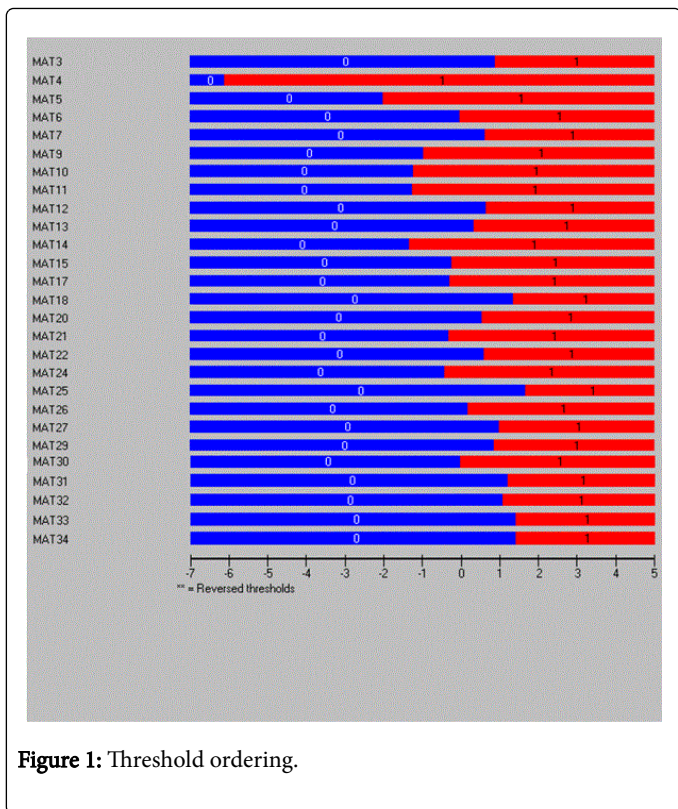
Figure 1: Threshold ordering.

Differential item functioning

DIF analysis for age and gender was conducted. The p values for all items exceeded the 0.000617 bonferonni level for both age and gender (Table 1). This indicates that given the same level of non-verbal intelligence, the expected score on any item was the same irrespective of the person factors (i.e. age, gender).

	Age				Age			
	Uniform DIF		**Non-uniform DIF**		**Uniform DIF**		**Non-uniform DIF**	
Items	F	P	F	P	F	p	F	p
MAT – 3	1.18	0.28	1.12	0.29	1.18	0.28	1.12	0.29
MAT – 4	1.56	0.21	2.96	0.09	1.56	0.21	2.96	0.09
MAT – 5	1.5	0.22	1.37	0.24	1.5	0.22	1.37	0.24
MAT – 6	3.02	0.08	2.09	0.15	3.02	0.08	2.09	0.15
MAT – 7	0	0.97	0.01	0.93	0	0.97	0.01	0.93
MAT – 9	0.18	0.67	1.1	0.29	0.18	0.67	1.1	0.29
MAT – 10	0.06	0.81	0.48	0.49	0.06	0.81	0.48	0.49
MAT – 11	0.01	0.93	0.17	0.68	0.01	0.93	0.17	0.68
MAT – 12	0.01	0.94	0.18	0.67	0.01	0.94	0.18	0.67
MAT – 13	4.89	0.03	-2.52	0.99	4.89	0.03	-2.52	0.99
MAT – 14	0.47	0.49	0.09	0.77	0.47	0.49	0.09	0.77
MAT – 15	0.6	0.44	-0.18	0.99	0.6	0.44	-0.18	0.99
MAT – 17	7.84	0.01	-4.83	0.99	7.84	0.01	-4.83	0.99
MAT – 18	0.12	0.73	0.24	0.62	0.12	0.73	0.24	0.62
MAT – 20	0.78	0.38	0.23	0.63	0.78	0.38	0.23	0.63
MAT – 21	0.39	0.53	-0.11	0.99	0.39	0.53	-0.11	0.99
MAT – 22	0.97	0.33	1.94	0.17	0.97	0.33	1.94	0.17
MAT – 24	2.21	0.14	6.09	0.01	2.21	0.14	6.09	0.01
MAT – 25	1.85	0.18	-0.58	0.99	1.85	0.18	-0.58	0.99
MAT – 26	0.01	0.9	0.07	0.78	0.01	0.9	0.07	0.78
MAT – 27	1.78	0.18	-0.04	0.99	1.78	0.18	-0.04	0.99
MAT – 29	0.8	0.37	1.25	0.26	0.8	0.37	1.25	0.26
MAT – 30	0.61	0.43	0.03	0.85	0.61	0.43	0.03	0.85
MAT – 31	1.42	0.24	1.44	0.23	1.42	0.24	1.44	0.23
MAT – 32	0.52	0.47	2.88	0.09	0.52	0.47	2.88	0.09
MAT – 33	3.46	0.06	0.66	0.42	3.46	0.06	0.66	0.42
MAT – 34	1.82	0.18	-1.1	0.99	1.82	0.18	-1.1	0.99

Table 1: DIF analysis for gender and age (N=102, df=1, Bonferonni adjustment α=0.000617).

Targeting and reliability

Figure 2 demonstrates the mapping of the distribution of persons' ability (upper part of the graph) and items' thresholds and their distribution (lower part) onto a horizontal scale indicating where the threshold for each item response was lying. The person location (M=-0.136) shows that on average children performed at a slightly lower level but within the acceptable range in non-verbal intelligence than the average of the scale items (M=0.00). This suggests that on the whole the scale was reasonably well targeted for the use with this random group of typically developing children. The thresholds at the left end are those which are the easiest to achieve and those at the right

end are the hardest to achieve. Generally there was an even spread of items across the full range of respondents' scores, suggesting effective targeting. Only two participants had higher ability than the difficulty of the existing items and there is some distance between the easiest item and the next item in line of difficulty.
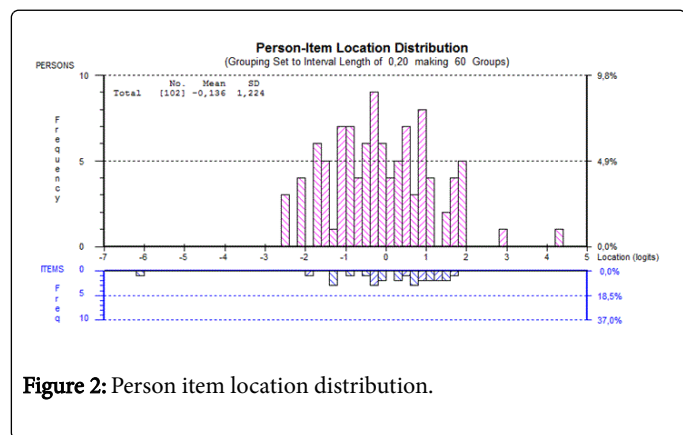

Figure 2: Person item location distribution.

The final PSI of 0.832 indicates that the MAT-SF has good person separation reliability and it can statistically discriminate between three groups of respondents.

Criterion validity

Criterion validity was tested by assessing the tool's ability to discriminate between children of different age groups and between higher and lower reading ability. Significant difference was identified for age (ANOVA; $F_{(1,101)}=13.53$; $p<0.0001$) with 9-11 years old children having greater ability (M=0.31) than 7-9 years old (M=-0.53). For reading ability, the children were divided in two groups based on their median score (<30 and >30). MAT-SF significantly discriminated these two groups of children (ANOVA; $F_{(1,101)}=43.82$; $p<0.0001$) with 9-11 years old children having greater ability (M=0.52) than the 7-9 years old (M=-0.83).

Discussion

The purpose of this study was to establish whether MAT-SF meets the formal requirements of measurement as defined by the Rasch model to be used for screening of non-verbal reasoning with Greek school-aged children of 7-11 years. There is only one previous study [22], which has partially tested for its validity with Greek population. However, further investigation of its validity and reliability was needed. Furthermore, this tool had not been previously Rasch evaluated. Seven misfitting items were identified and removed. The Rasch-scaled 27-item MAT-SF demonstrates a justifiable scale for screening non-verbal intelligence in Greek school-aged children. It possesses good reliability, demonstrated validity and effective targeting and shows no evidence of differential item functioning.

The seven items that were removed (I002, I008, I016, I019, I023, I028 and I035) to achieve fit to the overall model were outside the fit range of ± 2.5 or showed evidence of local dependency. Looking at the factor analysis of the initial validation of MAT-SF with US students, items I028 and I035 were also poorly loading on the first unrotated factor (i.e.<0.30) and I002 was problematic for some age groups [21]. It has been suggested that the variability in items outside the suggested range generates a substantial level of noise, which contributes little to

the measurement characteristics of the scale [39]. Therefore, it is important to remove such items.

Evidence of substantial construct validity of the Rasch-scaled revised MAT-SF is supported by the absence of DIF for age and gender. Also, on removal of items I008, I016 and I019, the test of local independence revealed no evidence of multidimensionality, suggesting that the tool is a unidimensional measure of non-verbal intelligence. Criterion validity was demonstrated by its ability to discriminate significantly between younger and older children and between children of higher and lower reading ability. Findings from other studies [7,21,40] confirm these results; thus indicating the importance of this test to screen Greek children's non-verbal intelligence. The categorisation of the items and its sensitivity for the ordering of the items thresholds were examined. There were no disordered thresholds indicating proper ordering.

The person-item distribution of the Rasch-scaled MAT-SF shows good targeting of the scale, with no apparent floor or ceiling effect. We found only two children who did not have difficulty performing even the most difficult items. However, the identified gap between the easiest item and the next item in line of difficulty suggests that future work should look at adding some items of low difficulty to cover for this gap. This could result in a more complete tool to assess non-verbal reasoning in Greek children with disabilities who might present difficulties in this area.

The Person-separation index (PSI) showed that the revised MAT-SF had the ability to discriminate between three groups of children, which indicated good reliability of the tool. It is interesting to note that both the original and the revised scales showed high values of the PSI, which corresponds to Cronbach's alpha. Actually, the original one showed slightly higher reliability index. Taking into account the necessary removal of items for the revised MAT-SF, the above indicates that measures of reliability do not always reflect quality of measurement. Instead they should be considered as strict measures of precision only if the items in the scale work invariantly [41]. Also during the original standardisation of MAT-SF with US students, the internal consistency of the tool was evaluated by using Cronbach's alpha. The PSI calculation through the Rasch analysis is a more appropriate method for assessing reliability. Although Cronbach's alpha is the most widely reported measure of internal consistency, the data need to meet certain assumptions for its use and it is dependent on the number of items included [42]. This is not the case with the PSI reports.

The limitations of this study relates to the age range of the sample. The initial test was validated with people 5-17 years, whereas the Greek sample in this study was 7-11 years. Future studies using Rasch analysis will need to be performed with other age groups in Greece to ensure comparability of item difficulties across ages. Also, future work should test Greek children with known disabilities, such as children with Down's syndrome [43] to see whether the tool has the sensitivity to differentiate children with disabilities from typically developing children. Furthermore, considering that the participants came from only one part of Greece, there is a need to validate the revised MAT-SF with participants from a variety of areas in Greece.

Following the initial results reported by Petrogiannis et al. [22], this study provided further evidence for the use of MAT-SF with Greek children. It demonstrated that the application of the Rasch measurement model supports the revised 27-item MAT-SF as a valid scale for measuring non-verbal reasoning of Greek children aged 7-11 years old. The revised MAT-SF has potentially great impact for school

psychologists, for school teachers and for researchers in Greece, where there is a great need for standardised tools to assess non-verbal ability [22]. This could provide important information in order to identify children with learning disabilities [21] or gifted children from disadvantaged backgrounds [27,44] and decide on further action. Thus, it could also facilitate targeting of provision of services to address giftedness in Greece or decision making as to whether there is need for applying training programmes to improve non-verbal ability [45]. This is important considering the links of non-verbal ability with working memory and academic achievement.

References

1. Horn J, Hofer SM (1992) Major abilities and development in the adult period. In: Stenberg R, Berg C (eds.) Intellectual development. Cambridge University Press, United Kingdom, pp. 44-91.

2. Horn J, Noll J (1997) Human cognitive capabilities: Gf-Gc theory. In: Flanagan D, Genshaft J, Harrison P (eds.) Contemporary intellectual assessment: Theories, tests and issues. Guilford Press, United Kingdom, pp. 53-91.

3. McArdle JJ, Ferrer-Caja E, Hamagami F, Woodcock RW (2002) Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. Dev Psychol 38: 115-142.

4. Godttfredson LS (1997) Why g matters: The complexity of everyday life. Intelligence 24: 79-132.

5. Alloway TP, Gathercole SE, Adams A, Willis C, Eaglen R, Lamont E (2005) Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. British Journal of Developmental Psychology 23: 417-426.

6. Lynn R, Meisenberg G, Mikk J, Williams A (2007) National IQs predict differences in scholastic achievement in 67 countries. J Biosoc Sci 39: 861-874.

7. Taub GE, Keith TZ, Floyd RG, Mcgrew KS (2008) Effects of general and broad cognitive abilities on mathematics achievement. School Psychology Quarterly 23: 187-198.

8. Floyd RG, Evans JJ, McGrew KS (2003) Relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. Psychology in the Schools 40: 155-177.

9. Mc Grew KS, Flanagan DP, Keith TZ, Vanderwood M (1997) Beyond g: The impact of Gf–Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. School Psychology Review 26: 189-210.

10. Fuchs LS, et al. (2006) The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. Journal of Educational Psychology 98: 29-43.

11. Swanson HL, Beebe-Frankenberge M (2004) The relationship between working memory and mathematical problem-solving in children at risk and not at risk for serious math difficulties. Journal of Educational Psychology 96: 471-491.

12. Gathercole SE, Alloway TP, Willis C, Adams AM (2006) Working memory in children with reading disabilities. J Exp Child Psychol 93: 265-281.

13. Floyd RG, Keith TZ, Taub GE, McGrew KS (2007) Cattell-Horn-Carroll cognitive abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effects. School Psychology Quarterly 22: 200-233.

14. Alloway TP, Gathercole SE, Pickering SJ (2006) Verbal and visuospatial short-term and working memory in children: are they separable? Child Dev 77: 1698-1716.

15. De Beni R, Palladino P, Pazzaglia F, Cornoldi C (1998) Increases in intrusion errors and working memory deficit of poor comprehenders. Q J Exp Psychol A 51: 305-320.

16. Gersten R, Jordan NC, Flojo JR (2005) Early identification and interventions for students with mathematics difficulties. J Learn Disabl 38: 293-304.

17. Swanson HL, Kehler P, Jerman O (2010) Working memory, strategy knowledge, and strategy instruction in children with reading disabilities. J Learn Disabl 43: 24-47.

18. Heitz RP (2006) Working memory, executive function, and general fluid intelligence are not the same. Behavioral and Brain Sciences 29: 135-136.

19. Bimey DP, Bowman DB, Pallier G (2006) Prior to paradigm integration, the task is to resolve construct definitions of gF and WM. Behavioral and Brain Sciences 29: 127.

20. Garlick D, Sejnowski TJ (2006). There is more to fluid intelligence than working memory capacity and executive function. Behavioral and Brain Sciences 29: 134-135.

21. Naglieri JA (1985) Matrix analogies test -Short form. U.S: The Psychological Corporation.

22. Petrogiannis KG, Bardos AN, Randou E (1999) Performance of Greek and American students on the Matrix analogies test: A measure of non-verbal intelligence. School Psychology International 20: 233-238.

23. Messick S (1995) Validity of psychological assessment:Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist 50: 741-749.

24. Nunnally J, Bernstein I (1994) Psychometric theory. New York: McGraw-Hill, Detroit, Michigan, United States.

25. Karabatsos G (2001) The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. J Appl Meas 2: 389-423.

26. Tafa E (1995) Test anixneusis tis anagnostikis ikanotitas. [Screening test for reading ability]. Ellinika, Grammata, Athens.

27. Prewett PN (1995) A comparison of two screening tests (the Matrix analogies test – Short form and the Kaufman brief intelligence test) with the WISC-III. Psychological Assessment 7: 69-72.

28. Howell KW, Zucker SH, Morehead MK (1985) Multilevel academic survey test. San Antoni, TX: The Psychological Corporation.

29. Naglieri JA (1985) Matrix analogies test – Expanded form. San Antonio, TX: The Psychological Corporation, Merrill.

30. Prewett PN, Farhney MR (1994) The concurrent validity of the Matrix analogies test—Short form with the Stanford-Binet: Fourth edition and KTEA-BF (Academic achievement). Psychology in the Schools 3: 20-25.

31. Prewett PN, Bardos AN, Naglieri JA (1988) Use of the MAT-SF and DAP: A quantitative scoring system with learning disabled and normal children. Journal of Psychoeducational Assessment 6: 347-353.

32. Bardos AN, Prewett PN (1991) Predictive validity of the Matrix analogies test - Short form. Diagnostique 16: 101-106.

33. Prewett PN, Bardos AN, Naglieri JA (1989) Assessment of mentally retarded children with the Matrix analogies test–short form, Draw–A–Person: A quantitative scoring system and the Kaufman test of educational achievement. Psychology in the Schools 26: 254-260.

34. Tennant A, Conaghan PG (2007) The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 57: 1358-1362.

35. Baghaei P (2008). The Rasch model as a construct validation tool. Rasch Measurement Transactions 22: 1145-1146.

36. Pallant JF, Tennant A (2007) An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 46: 1-18.

37. Smith EV Jr (2002) Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 3: 205-231.

38. Fisher W (1992) Reliability statistics. Rasch Measurement Transactions 6: 238.

39. Stelmack JA, Stelmack TR, Massof RW (2002) Measuring low-vision rehabilitation outcomes with the NEI VFQ-25. Invest Ophthalmol Vis Sci 43: 2859-2868.

40. Naglieri JA, Bardos AN (1988) Canadian children's performance on the Matrix analogies test. School Psychology International 9: 309-313.

41. Hagquist C, Bruce M, Gustavsson JP (2009) Using the Rasch model in nursing research: an introduction and illustrative example. Int J Nurs Stud 46: 380-393.

42. Spiliotopoulou G (2009) Reliability reconsidered: Cronbach's alpha and paediatric assessment in occupational therapy. Aust Occup Ther J 56: 150-155.

43. Hall P (2004) Effect of low doses of ionizing radiation in infancy on cognitive function in adulthood: Swedish population based cohort study. British Medical Journal 328: 19-23.

44. Alvino J, McDonnel R, Richert S (1981) National Survey of Identification Practices in Gifted and Talented Education. Exceptional Children 48: 124-132.

45. Bergman NS, Söderqvist S, Bryde S, Thorell LB, Humphreys K, et al. (2011) Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: a controlled, randomized study. Dev Sci 14: 591-601.