Journal of Psychoeducational Assessment
XX(X) xx–xx

DOI: 10.1177/0734282909343763
http://jpa.sagepub.com

Concurrent and Predictive Validity of the Raven Progressive Matrices and the Naglieri Nonverbal Ability Test

Giulia Balboni[1], Jack A. Naglieri[2], and Roberto Cubelli[3]

Abstract

The concurrent and predictive validities of the Naglieri Nonverbal Ability Test (NNAT) and Raven's Colored Progressive Matrices (CPM) were investigated in a large group of Italian third- and fifth-grade students with different sociocultural levels evaluated at the beginning and end of the school year. CPM and NNAT scores were related to math and reading comprehension tests. The CPM and NNAT concurrent and predictive validity coefficients were moderate, and there were no differences between concurrent and predictive validities with reading scores. The mutual incremental concurrent validity of the NNAT with the math test was significantly higher than the corresponding CPM validity. Finally, via step-down hierarchical regression analysis, the authors found that the predictive validity of both nonverbal tests is independent from the participants' sociocultural level. These two relatively simple nonverbal tests are commonly used to assess general ability; however, the present study shows that they can also provide useful information for predicting the academic performance of students with different sociocultural levels.

Keywords

nonverbal intelligence test, prediction of achievement, bias in prediction, NNAT, CPM

The correlation between general intelligence and academic scores has been reported in the literature to be about .55 to .60 (e.g., Naglieri & Bornstein, 2003; Neisser et al., 1996). Even if it is widely accepted that scholastic aptitude tests (e.g., the Scholastic Assessment Test offered by the College Board) are more effective at predicting academic outcomes (e.g., Anastasi & Urbina, 1997), intelligence tests are also regularly employed for predicting students' academic success (e.g., Gustafsson & Undheim, 1996; Spinath, Spinath, Harlaar, & Plomin, 2006). Nevertheless, the role of intelligence in the prediction of academic achievement is still disputed. Some authors (e.g., Jensen, 1998;

[1]University of Valle d'Aosta, Aosta, Italy
[2]George Mason University, Fairfax, VA
[3]University of Trento, Italy

Corresponding Author:
Giulia Balboni, University of Valle d'Aosta, Strada Cappucini 2/A, 11100 Aosta, Italy
Email: g.balboni@univda.it

Watkins, Lei, & Canivez, 2007) have argued that this relationship is probably because of the causal role of the general intelligence factor (*g*). In contrast, it has been suggested that either that relationship is because of the influence of education on the development of *g* (e.g., Ceci, 1992) or school achievement depends on basic cognitive abilities (e.g., Fagan, 2000; Luo, Thompson, & Detterman, 2003; Naglieri & Das, 2005). We favor the latter explanation. Despite these theoretical controversies, the fact that tests of general intelligence, which are used for a wide range of purposes (e.g., identifying strength and weaknesses, or diagnosing developmental disorders), are considered to be able to predict academic achievement (Lynn & Mikk, in press) makes the examination of their validity as potential indicators of later school achievement an important issue to be addressed.

Most studies of intelligence and achievement (e.g., Fergusson, Horwood, & Ridder, 2005; Jensen, 1998; Naglieri & Bornstein, 2003; Naglieri & Rojahn, 2004; Rushton, Skuy, & Fridjhon, 2003; Watkins et al., 2007) have examined concurrent validity (i.e., the relationship between the scores obtained on intelligence and achievement tests administered at the same time). However, predictive validity (i.e., the relationship between the scores obtained on intelligence tests administered at the beginning to predict achievement tests at the end of the school year) is particularly relevant for educators who wish to use these scores as a means of anticipating future performance, especially if specialized instruction is provided (Naglieri, Brulles, Lansdowne, 2009). Indeed, if the magnitude of the criterion validity correlation coefficient is high, students with a low intelligence test score should be provided additional instructional supports or possibility more in-depth evaluation so as to reduce the possibility of future failure.

Several studies have investigated the concurrent validity of the most widely used intelligence tests (e.g., Fergusson et al., 2005; Jensen, 1998; Naglieri & Bornstein, 2003; Naglieri & Rojahn, 2004; Rushton et al., 2003; Watkins et al., 2007). Concurrent validity correlation coefficients vary from moderate (.40 to .50) to high (.70 to .80) and appear to be independent of the tests used. Importantly, nonverbal tests of general intelligence have been found to be as predictive of achievement as those that also include verbal measures. Some authors (e.g., McCallum, 2003; Naglieri & Ford, 2003, 2005) have suggested, however, that nonverbal tests are more appropriate for assessing children with different levels of language, knowledge, and academic skills and with a low sociocultural background. The concurrent validity of nonverbal tests has been studied with different ethnic groups (e.g., Naglieri & Ronning, 2000a; Zucker, 1998); although rarely, the stability of the concurrent validity of intelligence tests has also been examined across ethnic groups (e.g., Weiss, Prifitera, & Roid, 1993; Young, 1994). Yet no studies have investigated the stability of the concurrent and predictive validities of intelligence tests across groups differing in sociocultural background.

The purpose of the present investigation was to examine the concurrent and predictive validities of two nonverbal intelligence tests relative to math and reading comprehension achievement scores in primary school children with different sociocultural levels. Raven's Colored Progressive Matrices (CPM; Raven, 1984) is one of the most employed instruments in both educational and clinical settings (e.g., Raven & Raven, 2003); the simple instructions and the low amount of time required make it a very suitable instrument. The CPM has good concurrent (Pind, Gunnarsdottir, & Johannesson, 2003; Rohde & Thompson, 2007) and predictive validity (Rushton et al., 2003) as well as split-half reliability (Raven & Raven, 2003). However, there is evidence that test–retest reliability of the CPM appears to be weak for intervals longer than 1 year (e.g., Kazlauskaite & Lynn, 2002; Raven & Raven, 2003). The Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997) has good concurrent validity (Lidz & Macrine, 2001; Naglieri & Ronning, 2000a, 2000b), even for different ethnic groups with comparable sociocultural levels (Naglieri & Ronning, 2000a). No studies, however, have assessed its predictive validity over time. NNAT is quite similar to the CPM, but given the different formats, the different numbers of items, and the different ways in which the items were constructed, standardized, and normed, some differences in validity are anticipated. In addition, no study has been conducted to investigate the mutual incremental validity, which is the amount of variance in achievement

Table 1. Characteristics of All Participants and the Three Subgroups With Different Family Cultural Status (FCS)

	All Participants ($N = 253$)	FCS		
		Low ($n = 86$; 34%)	Moderate ($n = 97$; 38%)	High ($n = 70$; 28%)
Age (year.month)[a]				
M	9.4	9.5	9.3	9.4
SD	1.1	1.0	1.1	1.1
Gender				
Female	130	38	55	37
Male	123	48	42	33
School grade				
3rd	126	39	53	34
5th	127	47	44	36

a. Measured at the beginning of the school year.

that each of these tests could explain over the other (e.g., Hunsley & Mayer, 2003). This information is helpful when determining the relative value of each test.

In the present study, we explore (a) whether both tests show differences between the concurrent and the corresponding predictive validity with achievement tests, (b) whether the concurrent and the predictive validities of both tests differentially predict math and reading comprehension, (c) which of the two nonverbal intelligence tests has higher incremental concurrent and predictive validity, and (d) whether the NNAT and CPM concurrent and predictive validity is independent from the sociocultural level of the participants.

Method

Participants

Potential participants were students who attended public schools in a large school district located in an urban area in northern Italy. Children in Grades 3 and 5 were selected to better represent the elementary school curriculum. The principals of six schools agreed to be involved in the present investigation, and nine classes in each grade were randomly selected. Parental permission was obtained for 301 students (86% of the total number of enrolled students). Of this group, 269 were at school all 8 days (4 in October and 4 in May) during which the data collection was conducted. A total of 17 students (6%) obtained a score more than 3.29 standard deviations over or under the mean in at least one test, and thus, according to instructions to identify outlier participants (see Tabachnick & Fidell, 2001), they were excluded from the study. The final sample size of 253 is described in Table 1. All 253 children resided in Italy; they were Italian native speakers, except for 6 of them (2%) who recently immigrated to Italy. In all, 4 students (1%) were identified as having learning disabilities.

Family cultural status (FCS; e.g., Coscarelli, Balboni, & Cubelli, 2007) was measured by means of the parents' education level and the Family Cultural Interest Scale (FCIS; Balboni, Bianchi, & Cubelli, 2003). FCIS provides a measure of what is referred to as cultural capital, which is the knowledge of cultural codes that are relevant for the society in which the individual lives (Lamont & Lareau, 1988). Examples are behaviors such as reading books, magazines, and newspapers, attending concerts, and visiting museums (Teachman, 1987). The calibration of parental education involved a series of steps. First, a 3-point parent education level variable was created. We summed the mother

and father education levels and divided the score distribution obtained into three levels based on the 33rd and 66th percentile scores. Next, a 4-point variable based on quartile scores was developed to describe family cultural capital by summing mother and father FCIS scores. Finally, the parents' education level and the family cultural capital variables were summed and the distribution of scores was divided into three categories based on the 33rd and 66th percentile scores. In this way, all participants were divided in three subgroups with low (34%), moderate (38%), and high (28%) FCS. Characteristics of the three subgroups are presented in Table 1. The three FCS subgroups did not differ regarding the gender, age, or school grade variables.

Instruments

CPM. The CPM is used for students from primary to secondary school and comprises 36 items organized in three different series. It is one of the most used nonverbal ability tests (Raven & Raven, 2003). In the present investigation, the recent Italian standardization of the CPM, which is based on 459 primary school children (Pruneti et al., 1996), was used.

NNAT. NNAT is a test of general ability for students from kindergarten through 12th grade. NNAT items are of a standard progressive matrix type ranging from 2 by 2 to 3 by 3 configurations with items that require the recognition of the patterns included in the matrix. The NNAT consists of seven levels each containing 38 dichotomously scored items. Each level contains some items from adjacent levels, which allows for continuous scaling of the test. A Nonverbal Ability Index standard score ($M = 100$, $SD = 15$) is converted from the child's NNAT raw score based on a U.S. standardization sample of approximately 89,000 students aged 5 to 18 years (Naglieri, 1997). No significant differences between the mean NNAT scores achieved by Italian students and the U.S. standardization sample were found (Balboni, Robusto, Cristante, & Naglieri, 2009).

Furthermore, by means of Rasch analysis, it has been found that NNAT items have no bias in evaluating individuals from different geographic areas, thus suggesting that this nonverbal measure of general ability may have utility across countries such as the United States and Italy.

Math school achievement test. The standardized Italian Elementary Math Test (Amoretti, Bazzini, Pesci, & Reggiani, 1994) was used. It is a paper-and-pencil test developed to measure mathematical skills taught at Italian primary schools. The test measures arithmetic (e.g., written calculation, writing Arabic numerals to dictation), geometric (e.g., knowledge of definitions and formulas), and logical skills (e.g., completing series). An example of an arithmetic item is, "The fraction 3/10 is equal to: a) 3.01; b) 310; c) 0.3; d) 0.03; e) 10.3." An example of geometric item is, "Draw two parallel lines." An example of logic item is, "Finish the sequence in accord to the rule on which it is based: 3 15 4 20 5 25 __ __ __ 35 __."

For each grade, two versions are available to be used at the beginning and at the end of the school year. The test is not timed and is group administered. The standardization is based on about 380 students for each grade, with a total N of 3,172. The math achievement test has high reliability (Cronbach's α average = .95; Amoretti et al., 1994) and validity (i.e., it can differentiate students with and without a disability in mathematics; e.g., Passolunghi & Siegel, 2004).

Reading comprehension test. The standardized Italian Memory and Transfer (MT) reading comprehension test was used (Cornoldi & Colpo, 1998). The MT test is composed of texts to be read and 10 to 15 multiple-choice questions to be answered; it requires, for example, that children make semantic and lexical inferences, discover the main idea of the text, and so forth. The test is organized into four levels per grade to measure reading comprehension skills: beginning, intermediate, end of school year, and a more difficult version for in-depth analysis of reading comprehension. The MT test is not timed, and generally it is group administered. The standardization is based on a sample of 600 students taken from Italy (Tressoldi, Lonciari, & Vio, 2000). Reliability is high (test-retest reliability = .90; Cornoldi & Colpo, 1998) and validity is good (see Carretti, Cornoldi, De Beni, & Romanò, 2005, for the discrimination between good and bad readers).

FCIS. This scale assesses the cultural interests and activities of both parents (Balboni et al., 2003). Questions include, for instance, the number of books and magazines read in 1 year, the number of foreign languages studied, cinema and theater shows seen, visits to museums and exhibitions, and participation in different kinds of groups and associations during free time. Each parent was scored on 10 multiple-choice items with a 5-point Likert-type scale. The alpha reliability of the scale is .72, and its validity had recently been supported by means of factorial analyses (Coscarelli, 2008).

Procedure

In October 2003, NNAT Levels D and E, CPM, the third- and fifth-grade math and reading comprehension tests Beginning School Year (BSY) were administered by a research assistant. The scale about the cultural interest of the family was self-administered by the participants' parents. In May 2004, the same examiner readministered the NNAT and CPM Levels D and E and administered the math test ending school year version and the reading comprehension in-depth version. In October and in May, all of the tests were administered, one each day, in a counterbalanced order, with NNAT and CPM the first or the second day and the math and the reading comprehension tests the third or the fourth day. The average interval between the first and last test administrations was 6.16 days ($SD = 2.37$, range = 4 to 14).

Data Analysis

Standard scores were based on the respective Italian standardization samples for all tests except the NNAT, for which only U.S. norms are available. To examine the concurrent and predictive validity of NNAT and CPM in predicting the achievement scores, Pearson correlation coefficients were calculated. For each intelligence test, concurrent correlation coefficients were compared to the corresponding predictive coefficients (e.g., NNAT math test concurrent vs. NNAT math test predictive coefficients). Moreover, concurrent as well as predictive correlation coefficients were compared within the different achievement tests (e.g., NNAT math test vs. NNAT reading comprehension test concurrent coefficients). For all of these comparisons, procedures according to Hotelling (1931), Williams (1959), and Steiger (1980) for two nonindependent correlation coefficients were used.

The mutual incremental concurrent and predictive validity of NNAT and CPM correlations with achievement test were investigated using linear hierarchical regression analysis. NNAT and CPM scores were entered as the first or as the second predictor of each achievement test score (criteria variable). Any statistically significant improvement in R^2 caused by the second predictor entered signifies that the second predictor has incremental validity compared to the first one. The magnitude of the incremental validity may be inferred in accordance with Cohen (1988): $\Delta R^2 = .02$ indicates a small and $\Delta R^2 = .13$ indicates a medium effect size. Moreover, to ascertain any statistically significance differences in the magnitude of the ΔR^2 pair of incremental validity (i.e., obtained with NNAT and CPM entered as first and second and then second and first predictors), the χ^2 test was used.

An analysis of the FCS predictive bias of ability tests in estimating achievement test scores was investigated. The relation between a predictor variable and a criterion variable can be systematically biased in two ways. The first is a systematic difference in the slope of the regression line between the predictor and criterion variables (e.g., Nunnally & Bernstein, 1994). A significant difference between participants with different FCS scores in the magnitude of the correlation coefficients between the ability and the achievement tests would indicate a bias in the accuracy of the prediction across the range of the predictor scores, generally labeled slope bias. A bias can also be found when the predictor variable systematically under- or overpredicts the criterion variable for a particular FCS group. Note that the degree of association between the predictor and the criterion variable need not differ across groups (e.g., Nunnally & Bernstein, 1994).

A standard means of investigating possible prediction bias and identifying slope and intercept differences is a step-down hierarchical multiple regression procedure described by Lautenschlager and Mendoza (1986; examples of application are available in Arbisi, Ben-Porath, & McNulty, 2002; Rotundo & Sackett, 1999). This method tests the hypothesis that one regression line can be used for all the different groups. If the hypothesis is rejected, this means that the prediction is significantly improved by using a separate regression line for each group.

First, an omnibus test of prediction bias was conducted using the regression model that included only the predictor variable (NNAT or CPM) compared to the full model that included the predictor variable (NNAT or CPM), the suspected moderator variable (FCS), and the cross-product of the predictor and the moderator variables. A significant increment in R^2 obtained by the use of the full model rather than the model containing the predictor alone signifies the presence of bias. To determine whether the bias is the result of differences in slope, intercept, or both, a series of tests for slope or intercept bias was performed. To test for slope bias, the full model was compared to a model containing NNAT or CPM and the FCS group variable (and not the cross-product of them). If a significant increment in R^2 is obtained, then this indicates the presence of slope bias, and a further test is performed to detect intercept bias. This test involves a comparison between the full model and a model containing NNAT or CPM and the cross-product of the FCS group variable and NNAT or CPM (and not the FCS group variable). If this comparison results in a significant increment in R^2, then intercept bias is also present; if there is no significant increase in R^2, then the bias identified is solely because of differences in slope. On the other hand, if the omnibus test for bias is significant but no slope bias is identified, then a separate test for intercept bias is performed comparing a model containing only NNAT or CPM to a model containing NNAT or CPM and the FCS group variable (and not the cross-product of the FCS group variable and NNAT or CPM). Again, if a significant increment in R^2 is obtained, then the presence of intercept bias is indicated.

Because the statistically significant slope and/or intercept bias may be small from a practical perspective (Lautenschlager & Mendoza, 1986), the standardized score differences were computed as effect sizes. To do so, the standardized score differences were computed by comparing the average predictive score obtained from the separate FCS group regression equation to the average predicted score that would be obtained from a common equation and dividing all by the standard deviation scores obtained from a common equation.

Results

Means, standard deviations, and skewness and kurtosis values for all measures at the beginning and at the end of the school year are presented in Table 2. The NNAT and CPM means at the end of the school year were higher ($p < .001$) than those obtained at the beginning, $t_{(252)} = 8.48$, $t_{(252)} = 9.22$, respectively. This is probably because of the effect of learning matrix items (e.g., Reeve & Lam, 2005). For both the CPM evaluation obtained in October and that in May, there was a ceiling effect, whereas there appeared to be no ceiling effects for the NNAT. The reading comprehension test score in May was lower than that found in October, $t_{(252)} = 16.47$, $p < .001$, and this is probably because of the reading comprehension test versions used: In October the beginning test version was used, whereas at the end of the year the in-depth test version, which is more difficult and allows for a more detailed evaluation, was used. We used the in-depth version of the reading comprehension test to avoid the ceiling effect that we found for the version administered in October.

Concurrent validity (measured both at the beginning and at the end of the school year) as well as predictive validity of Raven's and Naglieri's tests are shown in Table 3. Correlation coefficients were compared to verify the presence of any significant differences between the concurrent validity and the corresponding predictive validity coefficients obtained for the CPM and NNAT with each achievement test (e.g., NNAT math test concurrent vs. NNAT math test predictive coefficients). The

Table 2. Mean, Standard Deviation, Skewness, and Kurtosis of Naglieri Nonverbal Ability Test (NNAT), Raven's Colored Progressive Matrices (CPM), and Beginning School Year (BSY) and Ending School Year (ESY) Versions of the Achievement Test Standard Scores for All Participants

	Evaluation School Year Time							
	Beginning				Ending			
	M	*SD*	Skewness	Kurtosis	*M*	*SD*	Skewness	Kurtosis
NNAT	97.7	14.9	0.35	0.03	104.7	16.2	–0.04	0.22
CPM	103.2	14.7	–1.10	0.52	110.2	10.5	–1.32	1.51
Math BSY and ESY versions	100.9	14.4	–0.07	–0.82	100.5	14.8	–0.11	–0.63
Reading BSY and ESY versions	108.2	11.3	–1.22	0.94	94.6	15.7	–0.68	–0.20

Note: $N = 253$. At the beginning of the school year, NNAT Levels D and E, CPM, and reading and math BSY tests were used. At the end of the school year, NNAT Levels D and E and CPM were readministered and the in-depth version of the reading test as the ESY version and the ESY version of the math test were used.

Table 3. Concurrent and Predictive Correlation Coefficient Validities Among the Naglieri Nonverbal Ability Test (NNAT), Raven's Colored Progressive Matrices (CPM), and the Achievement Tests

	Concurrent Validity	Concurrent Validity Replication	Predictive Validity
NNAT—math	.52[a,b]	.45	.36[a]
CPM—math	.41[c]	.32	.33[c]
NNAT—reading	.40[b]	.44	.40
CPM—reading	.35	.40	.38

Note: $N = 253$. Coefficients with the same superscript letters differ significantly ($p < .05$, one-tailed). At the beginning of the school year, NNAT Levels D and E, CPM, and reading and math beginning school year tests were used. At the end of the school year, NNAT Levels D and E, CPM, the in-depth version of the reading test as the Ending School Year (ESY) version, and the math ESY tests were used. Concurrent validity and concurrent validity replication were obtained with beginning and end of school year data collection, respectively.

correlations suggest that, for both NNAT and CPM, the concurrent validity coefficient with the math test was significantly higher than the corresponding predictive validity coefficient, $t_{(250)} = 3.69$, $p < .001$; $t_{(250)} = 1.73$, $p < .05$, respectively. No significant differences were found when considering concurrent and predictive validity for the reading comprehension test. The presence of any significant differences between the corresponding criteria validity coefficients obtained for NNAT and for CPM with the two achievement tests (e.g., NNAT math test concurrent vs. NNAT reading comprehension test concurrent coefficients) was investigated. The NNAT math concurrent validity correlation coefficients obtained in October were significantly higher than the corresponding coefficients for the reading test, $t_{(250)} = 2.05$, $p < .05$. There were, however, no significant differences between the coefficients obtained with the math and reading comprehension tests for replication concurrent validity as well as for predictive validity. For CPM, no significant differences were found between the coefficients for the math and reading comprehension tests.

Hierarchical regression analysis, reported in Table 4, was used to investigate the mutual incremental validity of the NNAT and CPM in estimating the achievement scores in concurrent design (measured both at the beginning and at the end of the school year) and in predictive design (obtained comparing beginning and end of school year scores). Both the NNAT and the CPM had significant

Table 4. Regressions Predicting Math and Reading Comprehension Test Scores by Naglieri Nonverbal Ability Test (NNAT) and Raven's Colored Progressive Matrices (CPM) as First or Second Entered Predictor in Concurrent and Predictive Design: Comparison of the Corresponding ΔR^2

	NNAT Entered First, CPM Entered Second		CPM Entered First, NNAT Entered Second		
	R^2	ΔR^2	R^2	ΔR^2	χ^2
Concurrent validity					
Math BSY	.268***	.021**	.165***	**.124****	7.89**
Reading BSY	.163***	.025**	.124***	.063***	1.72
Concurrent validity replication					
Math ESY	.205***	.003	.104***	**.105****	10.18***
Reading ESY	.190***	.028**	.158***	.059***	1.15
Predictive validity					
Math ESY	.131***	.024**	.108***	.047***	0.77
Reading ESY	.157***	.039***	.147***	.049***	0.12

Note: BSY = beginning school year; ESY = ending school year; ΔR^2 represents the change in the proportion of the variance accounted for by the addition of the second predictor. Values in bold are approaching a medium effect size (Cohen, 1988). $\chi^2_{(1)}$ test was used to reveal if there were statistical differences among the corresponding ΔR^2 values. Concurrent validity and concurrent validity replication were obtained with beginning and ending school year data collection, respectively.
**$p < .01$. ***$p < .001$.

mutual incremental validity. The CPM concurrent and predictive incremental validity magnitude effect sizes were not statistically significant. In contrast, the NNAT incremental validity was always significant; moreover, concurrent incremental validity and its replication regarding math approached a medium effect size and was statistically higher than the corresponding CPM incremental validity.

A summary of the series of hierarchical regression analyses used to examine FCS prediction bias is reported in Table 5 for the math tests and in Table 6 for the reading comprehension tests. For the math test, there was evidence of intercept bias both for NNAT and CPM concurrent validity with data collected in May and for predictive validity; however, no bias was revealed for the concurrent validity data collected in October. For the reading comprehension test, evidence of intercept bias for NNAT and CPM concurrent validity was found in October and in May and for predictive validity. In any case, there was no evidence of slope bias. For all the intercept bias revealed, ΔR^2 values were always approaching a small size (Cohen, 1988). To examine the practical meaningfulness of the intercept bias, the standardized score differences were computed between the achievement predicted score averages obtained from a common regression equation and the three separate FCS group regression equations with the different intercepts. No statistically significant differences were found. In particular, the absolute value of the z statistic varied from 0.07 to 1.04 for the math test and from 0.06 to 0.98 for the reading comprehension test. In other words, the practical effect of the intercept bias was minimal as there were no differences in achievement scores estimated by the same regression equations for all of the FCS groups and by the three different equations with the three intercepts for all FCS groups.

Discussion

This investigation aimed to compare criterion validity of the NNAT and CPM to math and reading comprehension achievement test scores for Italian primary school children. The CPM, NNAT, and achievement tests were administered to the same participants at the beginning and at the end of the school year. In contrast to previous investigations (e.g., Jensen, 1998) and for the first time, the present study examined concurrent and predictive validity with the same sample.

Table 5. Summary of Step-Down Hierarchical Regressions Predicting Math Test Scores by Naglieri Nonverbal Ability Test (NNAT) and Raven's Colored Progressive Matrices (CPM) in Concurrent and Predictive Design

	NNAT		CPM	
	b	ΔR^2	*b*	ΔR^2
Concurrent validity				
Full model				
Nonverbal cognitive test	0.483***		0.202	
FCS	0.832		–8.028	
Interaction	0.005		0.094	
Prediction bias		.005		.014
Concurrent validity replication				
Full model				
Nonverbal cognitive test	0.350*		0.390	
FCS	0.636		0.882	
Interaction	0.021		0.027	
Prediction bias		.022*		.041**
Slope bias		.000		.000
Intercept bias		.022*		.041**
Predictive validity				
Full model				
Nonverbal cognitive test	0.281		0.185	
FCS	0.474		–3.149	
Interaction	0.027		0.062	
Prediction bias		.027*		.032**
Slope bias		.000		.002
Intercept bias		.027*		.030**

Note: FCS = family cultural status. The full model includes the nonverbal cognitive test, the FCS variable, and the interaction term. To investigate the prediction bias, only the nonverbal cognitive test model was compared to the full model. If there was evidence of prediction bias, then the kind of bias (slope and/or intercept) was checked. To investigate the slope bias, the nonverbal cognitive test plus the FCS group variable model was compared to the full model. If there was evidence of slope bias, then the intercept bias was investigated comparing the nonverbal cognitive test plus the interaction term model to the full model. If there was no evidence for slope bias, then the intercept bias was investigated comparing the nonverbal cognitive test model with the nonverbal cognitive test and FCS group variable model. *b* represents the unstandardized regression coefficient. ΔR^2 represents the change in the proportion of the variance accounted for by the addition of the model. FCS is coded 1 for low, 2 for medium, and 3 for high FCS participants. Concurrent validity and concurrent validity replication were obtained with beginning and ending school year data collection, respectively.
*$p < .05$. **$p < .01$. ***$p < .001$.T

The correlations between intelligence and math and reading comprehension achievement tests were moderate and varied between .32 and .52. These results are consistent with those of other investigations that found moderate to high correlation coefficients between intelligence and school achievement tests (e.g., Jensen, 1998; Naglieri & Bornstein, 2003; Watkins et al., 2007) but somewhat lower than those previously reported, for example, for the NNAT (Naglieri & Ronning, 2000b). The magnitude of these values in contrast to previous findings may be related to differences in the achievement tests studied. Consistent with the suggestions for evaluating the magnitude of a correlation coefficient (e.g., Cohen, Cohen, West, & Aiken, 2003), our results suggest that both NNAT and CPM have good predictive validity and may be useful in the early identification of students who risk developing learning problems, especially in reading comprehension.

When comparing the coefficients obtained for each intelligence test with the two achievement tests' scores, the concurrent coefficients for the NNAT with the math test scores were higher than the corresponding coefficients with the reading comprehension test. This finding is in agreement with

Table 6. Summary of Step-Down Hierarchical Regressions Predicting Reading Comprehension Test Scores by Naglieri Nonverbal Ability Test (NNAT) and Raven's Colored Progressive Matrices (CPM) in Concurrent and Predictive Design

	NNAT		CPM	
	b	ΔR^2	*b*	ΔR^2
Concurrent validity				
Full model				
Nonverbal cognitive test	0.366**		0.125	
FCS	5.807		–4.424	
Interaction	–0.037		0.066	
Prediction bias		.025*		.032**
Slope bias		.001		.004
Intercept bias		.023**		.027**
Concurrent validity replication				
Full model				
Nonverbal cognitive test	0.626***		0.731***	
FCS	14.921		13.079	
Interaction	–0.107		–0.077	
Prediction bias		.039**		.055**
Slope bias		.007		.002
Intercept bias		.033**		.053**
Predictive validity				
Full model				
Nonverbal cognitive test	0.451**		0.321	
FCS	6.737		0.662	
Interaction	–0.030		0.032	
Prediction bias		.036**		.039**
Slope bias		.001		.001
Intercept bias		.036***		.039***

Note: FCS = family cultural status. The full model includes the nonverbal cognitive test, the FCS variable, and the interaction term. To investigate the prediction bias, only the nonverbal cognitive test model was compared to the full model. If there was evidence of prediction bias, then the kind of bias (slope and/or intercept) was checked. To investigate the slope bias, the nonverbal cognitive test plus FCS group variable model was compared to the full model. If there was evidence of slope bias, then the intercept bias was investigated comparing the nonverbal cognitive test plus the interaction term model to the full model. If there was no evidence for slope bias, then the intercept bias was investigated comparing the nonverbal cognitive test model to the nonverbal cognitive test and FCS group variable model. *b* represents the unstandardized regression coefficient. ΔR^2 represents the change in the proportion of the variance accounted for by the addition of the model. FCS is coded 1 for low, 2 for medium, and 3 for high FCS participants. Concurrent validity and concurrent validity replication were obtained with beginning and ending school year data collection, respectively.
*$p < .05$. **$p < .01$. ***$p < .001$.

other studies (e.g., Colom & Flores-Mendoza, 2007; Spinath et al., 2006), but only for the results obtained at the beginning of the school year.

The concurrent and predictive validities of the NNAT were compared to those of the CPM by means of the measurement of the incremental mutual validity. The mutual incremental concurrent validity of the NNAT with the math test was significantly higher than the corresponding CPM validity. Indeed, the NNAT, compared to CPM, predicted an extra 10% of the math test score variance. Thus, it appeared that NNAT could be used instead of CPM to better predict math achievement scores in concurrent research designs. Furthermore, it is worth noting that different school grade versions are available for NNAT, thus allowing for an accurate evaluation with none of the ceiling effects found for the CPM.

Our results suggest that tests that assess general intelligence using nonverbal stimuli can be used as tools for predicting academic achievement. The fact that NNAT and CPM have good concurrent and predictive validity with achievement scores does not mean that these tests should be used in place of achievement tests for measuring present achievement skills but that they may provide information that is useful for predicting future academic achievement. It is worth noting that when measuring intelligence with a general ability test, information could also be obtained regarding what students have learned and how well they might learn in the future. We are not claiming that it is important to be able to predict future academic performance by means of intelligence tests alone but that information derived from intelligent tests can be considered along with other information when trying to anticipate a student's future performance. The moderate degree of the correlations between these two nonverbal tests with achievement implies that other constructs (e.g., scholastic motivation; Spinath et al., 2006) can also contribute a large portion of the variance in achievement scores. Future investigations are needed to reveal which other variables interact with general intelligence in explaining achievement performance. Additional research is needed to investigate the validity of NNAT and CPM with students with learning disabilities who often exhibit a discrepancy between intelligence and achievement scores. For these students, we do not predict a strong relationship between intelligence and achievement scores, either concurrent or predictive, because by definition the disability includes a discrepancy between these scores (American Psychiatric Association, 2000).

NNAT and CPM are nonverbal measures of general ability and therefore can be used for evaluating students in such a way as to reduce the role of language, knowledge, and academic skills on the measurement of ability (e.g., Naglieri & Ford, 2003; Naglieri & Prewett, 1990), making these tests appropriate across sociocultural backgrounds (Johnson, McGue, & Iacono, 2007). In our investigation, the step-down hierarchical multiple regression was significant for intercept bias for NNAT and CPM concurrent and predictive validities with math and reading comprehension tests. The practical effect of the intercept bias was negligible. However, as there were no differences in the achievement scores estimated by the same regression equation for all the FCS groups and by the three different equations with the three intercepts for each FCS group, both NNAT and CPM may be used to provide information about the future achievement scores regardless of the students' FCS. Further investigations should examine whether CPM and NNAT also have measurement invariance and what the relationship between measurement invariance and invariance in prediction is (Millsap, 2007).

Our investigation contributes to the overall knowledge of the field and increases the knowledge base relating to the use of nonverbal tests. However, this study has limitations that must be recognized. First, the reading comprehension achievement test used at the beginning of the school year showed a ceiling effect that was eliminated by using a longer version at the end of the school year. Future research should include more comprehensive tests of achievement at both concurrent and predictive phases of a study. Second, although the sample size was adequate for the statistics used in this study, larger, more diverse samples would provide greater stability for the findings. In particular, additional research is needed with older students. For high school students, for example, the validity coefficients could be higher because the IQ scores should be more reliable (e.g., Neisser et al., 1996). Third, larger samples would allow for additional comparisons, thus investigating the role of different variables such as gender and ethnicity. Finally, direct comparison of these findings to those for the U.S. standardization sample could be achieved by careful matching of Italian and American children on a number of critical demographic variables. Despite these limitations, we can conclude that a relatively simple nonverbal test of general ability, if supported by a careful analysis of its psychometric qualities, appears to be useful in an educational setting.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The investigation was partially supported by a grant from University of Valle d'Aosta. We are very grateful to Federica Bianchi for the excellent data collection.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

Amoretti, G., Bazzini, L., Pesci, A., & Reggiani, M. (1994). *Test di Matematica per la scuola dell'obbligo. Manuale* [Math school achievement test. Manual.]. Florence, Italy: Organizzazioni Speciali.

Anastasi, A., & Urbina, S. (1997). *Psychological testing.* Upper Saddle River, NJ: Prentice Hall.

Arbisi, P. A., Ben-Porath, Y., & McNulty, J. (2002). A comparison of MMPI-2 validity in African American and Caucasian psychiatric inpatients. *Psychological Assessment, 14,* 3-15.

Balboni, G., Bianchi, F., & Cubelli, R. (2003). *Questionario sugli interessi culturali* [Family Cultural Interest Scale]. Unpublished manuscript.

Balboni, G., Robusto, E., Cristante, F., & Naglieri, J. A. (2009). *Cultural bias in the Naglieri Nonverbal Ability Test*. Manuscript in preparation.

Carretti, B., Cornoldi, C., De Beni, R., & Romanò, M. (2005). Updating in working memory: A comparison of poor and good comprehenders. *Journal of Experimental Child Psychology, 91*, 45-66.

Ceci, S. J. (1992). Schooling and intelligence. *Psychological Science Agenda, 5,* 7-9.

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

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Colom, R., & Flores-Mendoza, C. E. (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence, 35,* 243-251.

Cornoldi, C., & Colpo, G. (1998). *Prove di Lettura MT per la scuola elementare-2. Manuale* [MT Reading Comprehension Test-2. Manual.]. Florence, Italy: Organizzazioni Speciali.

Coscarelli, A. (2008). *Misurazione del livello socioculturale in adulti: Differenze in test di abilità logiche e linguistiche* [Assessment of sociocultural level of adults: Differences in its effects in logical and linguistic tests]. Unpublished thesis, University of Aosta, Italy.

Coscarelli, A., Balboni, G., & Cubelli, R. (2007). The problem of measuring the socio-cultural level in psychological research. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities, volume 20: International perspectives* (pp. 163-181). Oxford, UK: Elsevier.

Fagan, J. F. (2000). A theory of intelligence as processing: Implications for society. *Psychology, Public Policy, and Law, 6,* 168-179.

Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2005). Show me the child at seven II: Childhood intelligence and later outcomes in adolescence and young adulthood. *Journal of Child Psychology and Psychiatry, 46,* 850-858.

Gustafsson, J.-E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186-242). New York: Prentice Hall.

Hotelling, H. (1931). The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. *Annals of Mathematical Statistics*, *11*, 271-283.

Hunsley, J., & Mayer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issue. *Psychological Assessment, 15,* 446-455.

Kazlauskaite, V., & Lynn, R. (2002). Two-year test-retest reliability of the Coloured Progressive Matrices. *Perceptual and Motor Skills, 95*, 354.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Johnson, W., McGue, M., & Iacono, W. G. (2007). Socioeconomic status and school grades: Placing their association in broader context in a sample of biological and adoptive families. *Intelligence, 35*, 526-541.

Lamont, M., & Lareau, A. (1988). Cultural capital: Allusions, gaps and glissandos in recent theoretical developments. *Sociological Theory, 6,* 153-168.

Lautenschlager, G. J., & Mendoza, J. L. (1986). A step-down hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement, 10,* 133-139.

Lidz, C. S., & Macrine, S. L. (2001). An alternative approach to the identification of gifted culturally and linguistically diverse learners. *School Psychology International*, *22,* 74-96.

Luo, D., Thompson, L. A., & Detterman, D. K. (2003). The causal factor underlying the correlation between psychometric *g* and scholastic performance. *Intelligence, 31,* 67-83.

Lynn, R., & Mikk, J. (in press). National IQs predict educational attainment in math, reading and science across 56 nations. *Intelligence*.

McCallum, R. S. (2003). Context for nonverbal assessment of intelligence and related abilities. In R. Steve & R. S. McCallum (Eds.), *Handbook of nonverbal assessment* (pp. 3-21). New York: Kluwer.

Millsap, R. E. (2007, July). *Invariance in measurement and prediction revisited*. Paper presented at the 72nd annual meeting of the Psychometric Society, Tokyo.

Naglieri, J. A. (1997). *Naglieri Nonverbal Ability Test*. San Antonio, TX: Psychological Corporation.

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment, 21,* 244-260.

Naglieri, J. A., Brulles, D., & Lansdowne, K. (2009). *Helping gifted children learn*. San Antonio, TX: Pearson.

Naglieri, J. A., & Das, J. P. (2005). Planning, attention, simultaneous, successive (PASS) theory: A revision of the concept of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 136-182). New York: Guilford.

Naglieri, J., & Ford, D. Y. (2003). Addressing under-representation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly, 47*, 155-160.

Naglieri, J. A., & Ford, D. Y. (2005). Increasing minority children's participation in gifted classes using the NNAT: A response to Lohman. *Gifted Child Quarterly, 49*, 29-36.

Naglieri, J. A., & Prewett, N. P. (1990). Nonverbal intelligence measures: A selected review of instruments and their use. In R. W. Kamphaus & C. R. Reynolds (Eds.), *Handbook of psychological and educational assessment of children: Volume 1, intelligence and achievement* (pp. 348-370). New York: Guilford.

Naglieri, J. A., & Rojahn, J. (2004). Construct validity of the PASS theory and CAS: Correlations with achievement. *Journal of Educational Psychology, 96,* 174-181.

Naglieri, J. A., & Ronning, M. E. (2000a). Comparison of White, African-American, Hispanic and Asian children on the Naglieri Nonverbal Ability Test. *Psychological Assessment*, *12*, 328-334.

Naglieri, J. A., & Ronning, M. E. (2000b). The relationship between general ability using the Naglieri Nonverbal Ability Test (NNAT) and Stanford Achievement Test (SAT) reading achievement. *Journal of Psychoeducational Assessment*, *18*, 230-239.

Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77-101.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory.* New York: McGraw-Hill.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 88,* 348-367.

Pind, J., Gunnarsdottir, E. K., & Johannesson, H. S. (2003). Raven's Standard Progressive Matrices: New school age norms and a study of the test's validity. *Personality and Individual Differences*, *34*, 375-386.

Pruneti, C. A., Fenu, A., Freschi, G., Rota, S., Cocci, D., Marchionni, M., et al. (1996). Aggiornamento della standardizzazione italiana del test delle Matrici Progressive Colorate di Raven (CPM) [Update of the Italian standardization of Raven's Colored Progressive Matrices]. *Bollettino di Psicologia Applicata, 217,* 51-57.

Raven, J. C. (1984). *Progressive Matrici Colore*. *Serie A, Ab, B*. *Manuale di istruzioni* [Italian version Raven Progressive Matrices]. Florence, Italy: Organizzazioni Speciali.

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In R. Steve & R. S. McCallum (Eds.), *Handbook of nonverbal assessment* (pp. 223-237). New York: Kluwer.

Reeve, C. L., & Lam, H. (2005). The psychometric paradox of practice effects due to retesting: Measurement invariance and stable ability estimates in the face of observed score changes. *Intelligence, 33,* 535-549.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35,* 83-92.

Rotundo, M., & Sackett, P. R. (1999). Effect of ethnic on conclusions regarding differential prediction in cognitive ability test. *Journal of Applied Psychology, 84,* 815-822.

Rushton, J. P., Skuy, M., & Fridjhon, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, East Indian, and White engineering students in South Africa. *Intelligence, 31,* 123-137.

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34,* 363-374.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245-251.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Teachman, J. (1978). Family background, educational resources and educational attainment. *American Sociological Review, 52,* 548-557.

Tressoldi, P. E., Lonciari, I., & Vio, C. (2000). Treatment of specific developmental reading disorders derived from single and dual route models. *Journal of Learning Disabilities, 33,* 278-285.

Watkins, M. W., Lei, P.-W., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence, 35,* 59-68.

Weiss, L. G., Prifitera, A., & Roid, G. (1993). The WISC-III and the fairness of predicting achievement across ethnic and gender groups. *Journal of Psychoeducational Assessment Monograph Series: Advances in Psychoeducational Assessment,* 35-42.

Williams, E. S. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society*, *Series B*, *21*, 396-399.

Young, J. W. (1994). Differential prediction of college grades by gender and ethnicity: A replication study. *Educational and Psychological Measurement, 54*, 1022-1029.

Zucker, R. (1998). Issues and trends in cultural-fair assessment. *Intervention in School and Clinic, 34*, 103-106.

Bios

Giulia Balboni, Ph.D. is Researcher of Psychometrics at the Faculty of Psychology at the University of Valle d'Aosta, Italy, member of the AAIDD ad hoc Committee on the development of the Diagnostic Adaptive Behavior Scale and the author of many papers on test and research methodology.

Jack A. Naglieri, Ph.D. is a Professor of Psychology at George Mason University, Senior Research Scientist at the Devereux Center for Resilient Children, Senior Editor of the Journal of Attention Disorders, and the author of many scholarly publications, books, and tests, including the Naglieri Nonverbal Ability Test.

Roberto Cubelli is Full Professor of General Psychology at the Faculty of Cognitive Sciences and Education at the University of Trento, Italy, Managing Editor of the international journal "Cortex", President of the Italian Association of Psychology and the author of many papers on cognitive psychology and clinical neuropsychology.