TPM®

EXPLORING THE POSSIBILITY TO USE SET I OF ADVANCED PROGRESSIVE MATRICES AS A SHORT FORM OF STANDARD PROGRESSIVE MATRICES

MATTEO CIANCALEONI
CATERINA PRIMI
FRANCESCA CHIESI
UNIVERSITY OF FIRENZE

The present paper aimed to explore the possibility of using the Advanced Progressive Matrices – Set I (APM-Set I) as a short form of Standard Progressive Matrices for sound assessment of general mental ability in a short time frame. Confirmatory factor analysis for dichotomous data attested the unidimensionality of APM-Set I, and then the Rasch model was applied. All items showed an acceptable fit, and the increasing order of items difficulty level was partially confirmed. Validity measures were also performed. Results supported the possibility of using APM-Set I as a short form of the Raven Matrices test.

Key words: Raven matrices; General intelligence; Short form; Rasch model; Unidimensionality.

Correspondence concerning this article should be addressed to Francesca Chiesi, Dipartimento di Psicologia, Università di Firenze, Via di San Salvi 12 – Padiglione 26, 50135 Firenze (FI), Italy. E-mail: francesca.chiesi@unifi.it

INTRODUCTION

The Standard Progressive Matrices (SPM; Raven, 1941), developed to measure general mental ability, is considered the purest measure of general mental ability (*g*) or fluid intelligence. The SPM is composed by five sets (A, B, C, D, and E) of 12 black-and-white patterns that have to be completed from a multiple choice set of alternatives, with the problems arranged in a progressively increasing level of difficulty within and across sets. This increasing items difficulty level is a central property because the validity of the SPM bears on learning from experience during the test (Raven, Raven, & Court, 2000). Another fundamental property of the scale is the unidimensionality (e.g., Raven & Fugard, 2008).

Despite SPM popularity, a potential limitation is that the original 60-item test administration time, often as long as 60 minutes, may be too long for some purposes. Such administration time makes it troublesome for both clinicians and researchers to administer other psychometric tests or experimental tasks during the same session. Moreover, this time length increases the influence of other variables, such as respondents' attention level, boredom, fatigue, and motivation (Smith, McCarthy, & Anderson, 2000). In the literature there are examples in which the length of SPM has been reduced by administering only some sets. For instance, in clinical settings, where the use of SPM is suggested to assess patients' general mental ability (e.g., Raven, 1941, 2000);

clinicians administer only sets A, B, C, and D (Caffarra, Vezzadini, Zonato, Copelli, & Venneri, 2003). Unfortunately, this four-set version cannot be used as a short form of SPM because, on the one hand, it represents only a partial reduction (from 60 to 48 items) and, on the other the most difficult items are excluded.

A different solution to reduce the length of the SPM was adopted by Nathaniel-James and colleagues (Nathaniel-James et al., 2004). In clinical settings, they recommended the use of the 12 items of the Advanced Progressive Matrices-Set I (APM; Raven, 1962) as a short form of SPM starting from the assumption that they covered all the intellectual processes and the full range of difficulties sampled by the SPM (Raven, 1962; Zaaiman, van der Flier, & Thijs, 2001), and, like SPM full-form, items were characterized by an increasing difficulty level (Raven, 1962). However, the psychometric properties of APM-Set I have not been investigated yet.

Taking into account that the main characteristics of the Raven Progressive Matrices are unidimensionality and increasing items difficulty level, the Rasch model (Rasch, 1960), within the Item Response Theory (IRT), is particularly suitable to investigate these properties.

First, it was developed to describe the relationship between the examinees' ability and the items difficulty level by postulating that such relationship can be described by a non-linear monotonically increasing function inside a unidimensional framework (Andrich, 1988; Hambleton, 1989, 1994; Rasch, 1977; Wright, 1977). That is, the fundamental assumption of the Rasch model is that items measure the same unidimensional latent construct and vary in their difficulty level, just as in the Raven Progressive Matrices.

Second, when fitting the Rasch model, it is the simplest model inside the IRT, and therefore the preferable one (Andrich, 2004), allowing to obtain invariant item difficulty estimates. So, in investigating the increasing items difficulty level of the Raven Matrices, it offers the simplest way to highlight the difficulty ranking.

Finally, whereas the Rasch model was proposed for examining dichotomous data, the satisfactory fit to data collected using multiple choice items supports the conclusion that these models are also suitable when items are not really dichotomous, as in the Raven Progressive Matrices, but the correct/incorrect dichotomy is obtained collapsing the options representing the wrong alternatives (Henning, 1989).

Some studies (Gallini, 1983; van der Ven & Ellis, 2000) applied IRT models to the study of SPM. Specifically, Gallini (1983) applied the Rasch model to data obtained administering SPM[1] to middle school students. He found a satisfactory fit for the model, and observed that the increasing order of items difficulty level was by and large confirmed with minor exceptions. Van der Ven and Ellis (2000) reported controversial results testing the unidimensionality by means of the Rasch model, which was separately applied to the five subsets. The Rasch model was not rejected for sets A, C, and D, whereas it was rejected for sets B and E. The authors suggested an interpretation of the fact that the items of these sets measure at least two different dimensions related to the items different characteristics. Nonetheless, further inspection of the factor plot showed that the emergence of a second factor could be considered as an artefact due to the skewness of the subset scores.

The purpose of the present work was to explore the possibility to use Set I of APM (Raven, 1962) as a short form of SPM (Raven, 1941), following Nathaniel-James et al. (2004). We aimed to assess the psychometric characteristics of that scale by administering it to a large sample of middle school students consistently with the assumption that the SPM can be adequately used to assess general mental ability in children and adolescents. SPM ability to discriminate among older adolescents and young adults appears instead to be weakened. Indeed, although the SPM were developed to assess general mental ability from early childhood to old age, the scores achieved by general population samples from many countries have been increasing over the years (Flynn, 1987; Raven, 2000), and the APM was recommended (Raven, 2000) as the means to assess general mental ability in adults and young adults in order to better discriminate higher-ability levels. Moreover, the sample was chosen in agreement with the age range of the Italian SPM standardization sample that is composed by adolescents from 11 to 17 years old (Valseschini & Del Ton, 1994).

In detail, the present work tested the unidimensionality of APM-Set I by confirmatory factor analysis for dichotomous data, and the Rasch model was applied in order to evaluate the increasing order of items difficulty level. Additionally, in order to provide a validity measure, the relation between fluid intelligence and working memory (WM) was taken into account. Among the researchers that have been studying this relationship, some argued that WM is so highly correlated with fluid intelligence that they could be deemed isomorphic (Kyllonen, 2002; Stauffer, Ree, & Carretta, 1996), some stated that these two constructs are barely linked to each other (Deary, 2000; Kline, 2000), while most claimed that WM and fluid intelligence are closely related but not identical (Ackerman, Beier, & Boyle, 2005; Yuan, Steedle, Shavelson, Alonzo, & Oppezzo, 2006, for a review; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Fry & Hale, 1996; Kane, Hambrick, & Conway, 2005). Using Raven's Standard Progressive Matrices, Fry and Hale (1996) found the impact of WM on fluid intelligence to be statistically significant (0.38); Engle et al. (1999) and Conway et al. (2002) pointed out that WM was correlated (respectively 0.49 and 0.60) with general fluid intelligence.

METHOD

Participants

Participants were 828 11-to-16 year old students attending middle schools in suburban districts in central Italy with a mean age of 12.71 years (*SD* = 1.01); the sample was gender homogenous: 431 students were male (52%). Two hundred and seventy-eight students were 6th graders (mean age = 11.71; *SD* = 0.56), 264 were 7th graders (mean age = 12.65; *SD* = 0.60), and 285 students were 8th graders (mean age = 13.73; *SD* = 0.61).

Measures

Set I of APM (Raven, 1962) is composed by 12 matrices of increasing difficulty level, and the items cover the range of difficulty of SPM (Raven, 1962; Zaaiman et al., 2001). These items are composed of a series of perceptual analytic reasoning problems, each in the form of a matrix. The problems involve both horizontal and vertical transformation: figures may increase or decrease in size, and elements may be added, subtracted, flipped, rotated, or may show other progressive changes in the pattern. In each case, the lower right corner of the matrix is missing and the participant's task is to determine which of eight possible alternatives fits into the missing space so that row and column rules are satisfied. Consistently with the long-forms of the Raven's Matrices Tests, three items — derived from Set A of the SPM (Raven, 1941) — were used as a practice test before completing APM-Set I.

Working memory was measured through the Digit Span scale of the WISC-III (Wechsler, 2006). Participants were read sequences of numbers and asked to repeat some of them forward and others backward. Series begin with two digits and keep increasing in length, with two trials at each length.

Procedure

Set I of APM was administered in the classroom respecting privacy. Each participant solved the task individually. The average administration time was 15 minutes. The Digit Span scale was then administered to a subsample composed by 118 students, who took about 5 to 10 minutes to complete the task.

RESULTS

The unidimensionality of the scale was tested with a Confirmatory Factor Analysis (CFA) for dichotomous data using Mplus 3.0 software (Muthén & Muthén, 2004) that implemented the *Weighted Least Squares Means and Variance adjusted* (WLSMV) estimation method. WLSMV uses weighted least square parameter estimates from the diagonal of the weight matrix. This method is recommended for categorical variables on the basis of simulation studies (e.g., Muthén & Muthén, 2004). Fit indices — $\chi^2(43) = 96.15$, $p < .001$; $\chi^2/df = 2.23$, CFI = .97, TLI = .98, RMSEA = .04 — indicated that a single-factor model adequately represents the structure of APM-Set I. Factor loadings (Table 1) were all significant, and the scale showed a satisfactory internal consistency.

Given that the unidimensionality assumption was met, the items fit statistics were calculated through Winsteps 3.59 software (Linacre, 2005) in order to test the fit between the items and the Rasch model. Infit values were reported as mean squares in the form of chi-square statistics divided by their degrees of freedom, their expected values was 1 and the suggested cut-offs to attest good fit were between 0.7 and 1.30 (Wright, Linacre, Gustafsson, & Martin-Loof, 1994). Results (Table 2) showed no misfitting items because mean square infit statistics ranged from .82 to 1.24, indicating that the empirical data met the model's requirements.

The difficulty level parameters were obtained employing two estimation methods implemented in Winsteps: *Normal Approximation Estimation Algorithm* (PROX; Cohen, 1979) and *Joint Maximum Likelihood Estimation* (JMLE; Wright & Panchapakesan, 1969). Initially all pa-

TABLE 1
Standardized factor loadings (all significant at $p < .001$),
explained variance and internal consistency of APM-Set I

Item	*F*1
1	.88
2	.72
3	.68
4	.66
5	.77
6	.68
7	.72
8	.58
9	.69
10	.55
11	.30
12	.59
Variance explained	44.2%
Rho*	.89

*This index (Bagozzi, 1994) is computed through the CFA estimated factor loadings and measurement residuals applying the following formula: $\rho = \frac{(\sum \lambda_j)^2}{(\sum \lambda_j)^2 + \sum \delta_j}$

TABLE 2
Item infit statistics

Item	Infit	Item	Infit
1	.81	7	.92
2	.99	8	.99
3	.95	9	.94
4	1.00	10	1.13
5	.89	11	1.24
6	.97	12	.96

rameter estimates were set to zero. The first phase of the estimation used the PROX method which was employed to obtain rough estimates. Then the PROX estimates became the starting values for JMLE which adjusted them, again by iterating through the data, in order to obtain the final JMLE estimates. When the convergence criteria were satisfied, the iterative process ceased and the final estimation was obtained. In order to evaluate item parameter estimation stability, the Winsteps item reliability index was used. Item reliability refers to the ability of the test to define a distinction hierarchy of items along the measured variable. This index is calculated dividing the true item variance by the observed item variance, and ranges from 0 to 1 with values higher than .90 indicating high reliability (Wrigth & Master, 1982). The items difficulty measures covered a range between –2.72 ± .18 and 3.81 ± .12 logits[2] (Table 3), and the reliability of these estimates was very high (.99) indicating a high expected stability of the ability estimates across samples. In detail, results showed that some items appeared to be out of order. More in details, item 4, item 7, and item 10 were easier in comparison with items coming first in the scale, and item 11, and not item 12, resulted to be the most difficult one.

TABLE 3
Item difficulty level estimates with related standard errors (out of order items are in italics)

Item	Difficulty	Item	Difficulty
1	–2.72 ± .18	7	*–.60 ± .10*
2	–1.90 ± .14	8	1.12 ± .09
3	–.82 ± .11	9	1.63 ± .09
4	*–1.91 ± .14*	10	*.07 ± .09*
5	–.45 ± .10	11	*3.81 ± .12*
6	–.25 ± .10	12	2.01 ± .09

Pearson product-moment correlations relating APM-Set I score and Digit Span scores (*backward*, *forward*, *total*) were calculated. The results attested that all the relations were significant (*Span back*: $r(N = 118) = .22, p < .01$; *Span forward*: $r(N = 118) = .29, p < .001$; *Span Total*: $r(N = 118) = .30, p < .001$).

DISCUSSION

The present study aims to explore the possibility of using Set I of Advanced Progressive Matrices (Raven, 1962) as a short form of Standard Progressive Matrices (Raven, 1941) in order to avoid the potential limitations of the original 60-item test due to long administration time (e.g., the influence of boredom or tiredness). APM-Set I was previously used as a short form of the SPM in clinical contexts (Nathaniel-James et al., 2004) but its psychometric properties have not been investigated yet. In order to do that, the Rasch model for dichotomous data was applied to APM-Set I after attesting its one-factor structure through confirmatory factor analysis.

The Rasch model fitted the data, and the difficulty parameter estimates indicated that the items were distributed along the ability scale. The original ranking of item difficulty was only partially confirmed. Three items are easier than some others coming earlier in the scale, and in order to obtain an increasing difficulty level, the fourth item should be placed in the second position; the seventh item should be placed in the fifth, and the tenth item in the eighth. One item is harder than expected, suggesting that the two last positions should be inverted because the next-to-last, and not the last, item appears to be the most difficult one.

Finally, a validity index was performed. Scores obtained administering APM-Set I are positively correlated with working memory, as well as with the measure of general intelligence derived from the SPM (Conway et al., 2002; Engle et al., 1999; Fry & Hale, 1996).

In sum, the present results support the possibility of using APM-Set I as a short form of SPM. This characteristic could make the scale an efficient tool for researchers and practitioners to use for clinic or research purposes, especially when Raven's Matrices have to be administered along with other tests. Indeed, just like the long form, APM-Set I is a one-factor scale with items characterized by different difficulty levels — whereas their ranking is not totally consistent with the way in which items are supposed to increase in difficulty.

The former finding suggests that the items' ranking should be partially modified following the ranking order empirically identified in this study. Indeed, when interested in using APM-Set I as a short form of the Raven's Progressive Matrices, the suggested order better reproduces the main characteristic — the item difficulty increase — of the test.

Further studies may be needed to confirm and extend these findings. First, future investigations should be conducted with younger children in order to generalize APM-Set I use as a short form of SPM. Second, its validity should be further investigated in order to strengthen the possibility of using APM-Set I to provide a sound assessment of general intelligence in a short time frame.

NOTES

1. More precisely, the analysis was conducted administering only 50% of the SPM items. Nonetheless, these items were selected to insure that the overall difficulty pattern of the total test was maintained.
2. The logit is the logarithm of the *odd*, i.e., the ratio between the probability of the correct answer, $p(X = 1)$, and the probability of the wrong answer, $p(X = 0)$.

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30-60.
Andrich, D. (1988). *Rasch model for measurement*. London: Sage.
Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical care*, *42*, 17-116.
Bagozzi, R. P. (1994). Structural equation models in marketing research: Basic principles. In R. P. Bagozzi (Ed.), *Principles of Marketing Research* (pp. 317-85). Malden, MA: Blackwell.
Caffarra, P., Vezzadini, G., Zonato, F., Copelli, S., & Venneri, A. (2003). A normative study of a shorter version of Raven's progressive matrices 1938. *Journal of Neurological Sciences*, *24*, 336-339.
Cohen, L. (1979). Approximate expressions for parameter estimates in the Rasch model. *British Journal of Mathematical and Statistical Psychology*, *32*(1), 113-120.
Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163-183.
Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. New York: Oxford University Press.
Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309-331.
Flynn, J. R. (1987). Massive IQ in 14 nations: What IQ tests really measure? *Psychological Bulletin*, *101*, 171-191.
Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*, 237-241.
Gallini, J. K. (1983). A Rasch analysis of Raven item data. *Journal of Experimental Education*, *52*(1), 27-32.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 143-200). New York: Macmillan.
Hambleton, R. K. (1994). Item response theory: A broad psychometric framework for measurement advances. *Psicothema*, *6*(3), 535-556.
Henning, D. (1989). Does the Rasch model really work for multiple choice item? *Journal of Educational Measurement*, *26*(1), 91-97.
Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2004). *Psychological Bulletin*, *131*, 66-71.
Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). New York: Routledge.
Kyllonen, P. C. (2002). g: Knowledge, speed, strategies, or working memory capacity? A systems perspective. In R. J. Sternberg & E. L. Gigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 415-445). Mahwah, NJ: Erlbaum.
Linacre, J. M. (2005). *Winsteps* (Version 3.59) [Computer Software]. Chicago: Winsteps.com.
Muthén, L. K., & Muthén, B. O. (2004). *Mplus: The comprehensive modeling program for applied researchers. User's guide* (3rd ed.). Los Angeles: Muthén & Muthén.
Nathaniel-James, D. A., Brown, R. G., Maier, M., Mellers, J., Toone, B., & Ron, M. A. (2004). Cognitive abnormalities in schizophrenia and schizophrenia-like psychosis of epilepsy. *Journal of the Neuropsychiatry and Clinical Neurosciences*, *16*, 472-479.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, *14*, 58-94,
Raven, J. C. (1941). Standardization of progressive matrices. *British Journal of Medical Psychology*, *19*, 137-150.
Raven, J. C. (1962). *Advanced progressive matrices*. London: Lewis & Co. Ltd.
Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*, 1-48.
Raven, J., & Fugard, A. (2008). What's wrong with Factor-Analyzing Tests conforming to the requirements of Item Response Theory? *WebPsychEmpiricist*. http://wpe.info/papers_table.html.
Raven, J., Raven, J. C., & Court, J. H. (2000). *Raven manual: Section 3, standard progressive matrices, including the parallel and plus version.* Oxford: Oxford Psychologists Press.
Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short form development. *Psychological Assessment*, *12*(1), 102-111.
Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than g: An extension of Kyllonen's analyses. *The Journal of General Psychology*, *123*, 193-205.
Valseschini, S., & Del Ton, F. (1994). Le matrici progressive di Raven: contributo critico alla taratura su 1123 soggetti e considerazioni sulla validità e attendibilità della prova [Raven's Standard Progressive Matrices: a critical contribution to adjustment on 1123 subjects and considerations on the test validity and reliability]. Firenze, Italy: Organizzazioni Speciali.
van der Ven, A. H. G. S., & Ellis, J. L. (2000). A Rasch analysis of Raven's Standard Progressive Matrices. *Personality and Individual Differences*, *29*, 45-64.
Wechsler, D. (2006). *WISC-III. Wechsler Intelligence Scale for Children — Terza Edizione — Manuale* [*WISC-III. Wechsler Intelligence Scale for Children — Third Edition — Manual*]. Firenze, Italy: Organizzazioni Speciali.
Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, *14*(2), 97-116.
Wright, B. D., Linacre, J. M., Gustafsson, J-E., & Martin-Loff, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.
Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free items analysis. *Educational and Psychological Measurement*, *29*, 23-48.
Yuan, K., Steedle, J., Shavelson, R., Alonzo, A., & Oppezzo, M. (2006). Working memory, fluid intelligence, and science learning. *Educational Research Review*, *1*, 83-98.
Zaaiman, H., van der Flier, H., & Thijs, G. D. (2001). Dynamic testing in selection for an educational programme: Assessing South African performance on the Raven Progressive Matrices. *International Journal of Selection and Assessment*, *9*(3), 258-269.