Intelligence 36 (2008) 702–710

Contents lists available at ScienceDirect

Intelligence

The quest for item types based on information processing: An analysis of Raven's Advanced Progressive Matrices, with a consideration of gender differences ☆

François Vigneau [a,*], Douglas A. Bors [b]

[a] École de psychologie, Université de Moncton, Moncton, New Brunswick, Canada
[b] Department of Psychology, University of Toronto at Scarborough, Scarborough, Ontario, Canada

A R T I C L E I N F O

Article history:
Received 1 June 2007
Received in revised form 23 April 2008
Accepted 29 April 2008
Available online 9 June 2008

Keywords:
Intelligence
Raven's Progressive Matrices
Gender differences
Information processing
Dimensionality

A B S T R A C T

Various taxonomies of Raven's Advanced Progressive Matrices (APM) items have been proposed in the literature to account for performance on the test. In the present article, three such taxonomies based on information processing, namely Carpenter, Just and Shell's [Carpenter, P.A., Just, M.A., & Shell, P., (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. Psychological Review, 97, 404–431.] as completed by Mackintosh and Bennett [Mackintosh, N.J., & Bennett, E.S., (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. Intelligence, 33, 663–674.], DeShon, Chan and Weissbein's [DeShon, R.P., Chan, D., & Weissbein, D.A., (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. Intelligence, 21, 135–155.], and Dillon, Pohlmann and Lohman's [Dillon, R.F., Pohlmann, J.T., & Lohman, D.F., (1981). A factor analysis of Raven's Advanced Progressive Matrices freed of difficulty factors. Educational and Psychological Measurement, 41, 1295–1302.], were examined to assess the extent to which they fit APM data from two large samples of university students. Gender differences hypotheses based on the examined APM item taxonomies were also tested. Results indicate that none of the tested models achieved a good fit, and that item difficulty seemed to be the main determinant of the APM's dimensionality. Gender differences analyses also provided inconsistent support for the information-processing based taxonomies. Results are discussed in terms of potential statistical artifacts and of the weak reliability of the proposed item classifications.

Mackintosh and Bennett (2005) reported gender differences evidence supporting the taxonomy of information processing rules proposed by Carpenter, Just and Shell (1990) for solving the Raven's Advanced Progressive Matrices (APM; Raven, 1962) test items. They reported finding gender differences, in favour of men, on items requiring an "addition/subtraction" rule and on items requiring a "distribution of two" rule, but not on items requiring a "pairwise progression" rule or a "distribution of three" rule. Such a pattern of results supports a general distinction of APM items based on the type of information processing rules that they require to be solved. More specifically, this pattern of gender differences is also roughly consistent with a distinction made by DeShon, Chan and Weissbein (1995) between visuospatial and verbal–analytical items. In light of DeShon et al.'s (1995) findings, Mackintosh and Bennett (2005) interpreted their own findings of men outperforming women on specific items as being a consequence of these items' inclusion of a spatial component, a dimension of ability that has frequently been found to manifest a male advantage (Voyer, Voyer, & Bryden, 1995).

Mackintosh and Bennett's (2005) results are consistent with a number of dimensionality studies of Raven's Matrices

☆ This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the New Brunswick Innovation Foundation (NBIF).

* Corresponding author.
E-mail address: f.vigneau@umoncton.ca (F. Vigneau).

0160-2896/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2008.04.004

reporting multifactorial solutions. Most of these studies pertained to the Standard version of the Matrices (SPM, Raven, 1956; e.g., Lynn, Allik & Irwing, 2004; van der Ven & Ellis, 2000), but there have been reports indicating that the advanced version of the Matrices (APM) could also be multidimensional (Dillon, Pohlmann & Lohman, 1981).

The notion that Raven's Matrices items can be classified in terms of the information processing rules required for their solution was most notably developed by Carpenter et al. (1990). According to their proposed taxonomy, five rules are used to solve APM items. Keeping in mind that APM items are figure patterns of 9 cells arranged in 3 (rows)×3 (columns) matrices, these five rules can be described as follows:

C	"Constant in a row"	The same value of an attribute appears in all three cells in a row, but changes between rows.
PP	"Quantitative pairwise progression"	A constant change occurs in the size, number or position of an attribute between neighbouring cells in a row.
A/S	"Figure addition or subtraction"	A figure from one cell is added to (juxtaposed or superimposed), or subtracted from, a figure in a second cell to produce a figure in a third cell.
D3	"Distribution of three values"	A different value of a categorical attribute appears in each of the three cells of a row.
D2	"Distribution of two values"	Two values of a categorical attribute are distributed through the row, with the value for the third cell being null.

According to Carpenter et al. (1990), each APM item is characterized by one or two specific types of rule (often in addition to the constant-in-a-row rule). Such a categorization of APM items according to the proposed rules is not straightforward, however. The taxonomy originally offered (see Appendix A) was the result of several specific choices and decisions. Carpenter et al. (1990) derived their rule taxonomy from verbalization and eye-movement APM data collected from a sample of 12 student participants. These subjects were asked to verbalize their thoughts while solving APM items. The analyses leading to the rule taxonomy was row-oriented only (it did not take into account vertical or diagonal analysis). In cases where row-oriented analysis allowed more than one possible rule, resolution was in favour of the alternative verbalized by the "highest scoring subjects" in the sample. Furthermore, 11 of the 36 Set II items were excluded from the taxonomy—nine items due to technical difficulties, and two others because they were found to be unclassifiable in terms of the taxonomy's categories. Taken together, these specific research strategies and choices excluded a priori the discovery of alternative rules for solving any given item. It should be acknowledged, however, that the identification of the diversity of the rules and procedures that could be used successfully to solve any APM item might not have been Carpenter et al.'s (1990) goal. Mackintosh and Bennett's (2005, Table 1) own completion of the taxonomy for most of the 11 items left unclassified by Carpenter et al. (1990) also is subject to the same restrictions.

Another processing rule taxonomy of APM items arose from the experimental work of DeShon et al. (1995). Using the verbal overshadowing paradigm, they hypothesized that one subset of APM items were dependent on visuospatial processes, whereas another subset of items required verbal–analytic processes. They assumed that, because of limits inherent in information processing capacity, performance on visual items should be affected by concurrent verbalization, whereas verbal items should not. DeShon et al. (1995) found that the average performance on seven of the 12 items classified as visuospatial clearly suffered as a result of concurrent verbalization (the other five visuospatial items exhibited non-significant performance reductions in the verbalization condition), whereas none of the nine items classified as verbal–analytic showed a negative effect as a result of verbalization.

According to Mackintosh and Bennett (2005), DeShon et al.'s (1995) analysis follows Carpenter et al.'s (1990) closely: "all items requiring the distribution-of-three rule for their solution are classified as verbal–analytic by DeShon et al. (1995), while most of those requiring addition/subtraction or distribution of two rules are classified as visuospatial" (p. 665). The development of DeShon et al.'s (1995) taxonomy, however, was derived with even less data than was Carpenter et al.'s (1990). The classification that they produced (visual and analytic items; see Appendix A) was based on the independent coding of all APM Set II items by each of the three co-authors, with a reported average percent agreement for dominant processing mode (visuospatial or verbal–analytic) "near 100%". The objective of this classification was to characterize items as much as possible in terms of a dominant mode of processing. Presumably, this would again have left little room for a diversity of problem solving

Table 1
Goodness-of-fit measures of five measurement models of the APM (N=506)

Model	χ^2 (*df*)	GFI	CFI	NNFI	SRMR	AIC	SBC	RMSEA (90% CI)
Single-factor model								
1-factor (36 items)	148 (594)	.87	.75	.73	.055	−40	−2550	.043 (.039–.047)
Dillon et al.'s (1981) (36 items)								
2-factor (r=0.95)	1143 (592)	.87	.75	.73	.055	−41	−2543	.043 (.039–.047)
DeShon et al.'s (1995) *(25 items)*								
1-factor	581 (275)	.91	.79	.77	.054	31	−1131	.047 (.042–.052)
2-factor (r=1.00)	581 (274)	.91	.79	.76	.054	33	−1125	.047 (.042–.053)
Mackintosh and Bennett's (2005) *(33 items)*								
1-factor	981 (495)	.88	.75	.73	.056	−9	−2101	.044 (.040–.048)
4-factor (r=0.86)	958 (494)	.88	.76	.75	.056	−30	−2118	.043 (.039–.047)
Skewness-based 2-factor model (36 items)								
2-factor (r=0.58)	940 (593)	.90	.84	.83	.050	−246	−2753	.034 (.030–.038)

Note. GFI = goodness-of-fit index; CFI = comparative fit index; NNFI = non-normed fit index; SRMR = standardized root mean square residual; AIC = Akaike's information criterion; SBC = Schwarz's Bayesian criterion; RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval for RMSEA. Dillon et al.'s models: one pattern-progression factor (18 items) and one figure-addition/subtraction factor (18 items, including item 31). DeShon et al.'s models: one analytic factor (12 items) and one visual factor (13 items). Mackintosh and Bennett's models: 9 PP items, 8 A/S items, 9 D3 items, and 7 D2 items. Skewness-based model: one negative (items 1 through 23) and one positive (items 24 through 36) skewness factors. See Appendix for exact item assignations.

strategies to be taken into account. Again, the idea of multiple paths for arriving at the correct answer was apparently not of interest to the authors. In any case, several of the 36 Set II APM items could not be unequivocally classified into one of the dominant processing categories—four items were classified as "either" (equally likely to be solved using visuospatial or verbal–analytic strategies), six were classified as "both" (typically solved using visuospatial strategies, but also requiring verbal–analytic processing), and one was deemed unclassifiable (see Appendix A).

Interestingly, DeShon et al. (1995) did not obtain factor analytic results consistent with their experimental findings. Indeed, when they used confirmatory factor analysis to compare a correlated two-factor model based on their taxonomy to a single-factor model, they failed to obtain a better fit for the two-factor model. From this they concluded that their data were consistent with previous findings of Arthur and Woehr (1993) indicating, contrary to their taxonomy, that a single factor appears to be responsible for the observed inter-item correlations.

Although the evidence supporting a multidimensional view of the SPM (Kubinger, Formann, Farkas, 1991, Lynn, Allik & Irwing, 2004; van der Ven & Ellis, 2000), or even of Set I of the APM (Hunt, 1974), has accumulated over the years, no such clear picture has emerged in the case of Set II of the APM. Dillon et al. (1981) obtained results that indicated that the APM was dominated by two factors: one factor was interpreted as reflecting the ability to add and subtract patterns (addition/subtraction), and one factor was interpreted as the ability to detect a progression in a pattern (pattern progression) (see Appendix A). As mentioned above, however, Arthur and Woehr (1993) and DeShon et al. (1995) both failed to replicate Dillon et al.'s (1981) results. Abad, Colom, Rebollo and Escorial (2004), Alderton and Larson (1990), and Bors and Stokes (1998) also provided factor analysis results supporting a one-factor model of APM (Set II) items.

Arguing that, in the case of the APM, the acceptance of a one-factor solution is more of a default acquiescence than a compelling conclusion, Vigneau and Bors (2005) indicated that exploratory factor analyses of Bors and Stokes' (1998) data as well as of their new 12-item APM data revealed factors based on item difficulty, or item position in the test. Specifically, in a two-factor solution, items positioned early in the test generally clustered together, whereas items positioned later in the test formed their own separate cluster. Similarly, a three-factor solution illustrated a pattern of beginning-item, middle-item, and end-item clusterings. Such factor solutions based on item position in the test are difficult to interpret, however. Potential sources of item segregation include item position in the test, with the associated potential effects of learning and fatigue, item difficulty, and a possible artifact of the wide range in item skewnesses. Barring any experimental manipulation, these three potential sources of item separation remain confounded. Considering that Dillon et al. (1981) did not obtain factors based on item position in the test and used an untimed administration of the APM, whereas Vigneau and Bors (2005) did obtain such a pattern under timed conditions, it is also possible that item-position-based factor solutions reflect the common use of a time limit when administering the APM.

Mackintosh and Bennett (2005), using both published data and a specially designed experiment, reported two sets of results showing that gender differences on the APM were mainly associated with item clusters characterized as requiring visuospatial processing—namely, in Carpenter et al.'s (1990) terminology, the figure-addition/subtraction rule and the distribution-of-two-values rule. However, and despite their careful control of a potential confounding variable (item difficulty), Mackintosh and Bennett's (2005) results were not entirely consistent with the idea that the male advantage on Raven items was due to gender-related differences in visuospatial ability. Specifically, there was no indication in their data that scores on figure-addition/subtraction or distribution-of-two-values items were more strongly correlated with Vandenberg and Kuse's (1978) Mental Rotation Test, a measure of visuospatial ability, than were scores on other items, as might be expected. The reliability of Macintosh and Bennett's (2005) findings has also been questioned recently by Colom and Abad (2007), who were unable to replicate, in a large sample of 1970 adults, Mackintosh and Bennett's (2005) findings of a specific relation of gender to figure-addition/subtraction and distribution-of-two-values items. Rather, Colom and Abad (2007) observed significant associations of small magnitude between gender and all categories of item types. These authors also pointed out that the gender effect reported by Mackintosh and Bennett (2005) (d=43) was considerably higher than that typically found in other published studies.

The present research addresses the notion that the items of the advanced version of Raven's matrices (APM, Set II) can reliably be classified in terms of information processing rules required for their solution. To examine this question, two versions of the APM—the original 1962 revision, and a specially designed 14-item short-form based on DeShon et al.'s (1995) proposed taxonomy and results—were used in two distinct samples of subjects. In Study 1, previously published 36-item APM data from a large sample of university students (Bors & Stokes, 1998; Vigneau & Bors, 2005) were reanalyzed. In Study 2, new data collected with a 14-item short version of the APM (VA-APM; Vigneau & Bors, 2001) were used. Three item taxonomies suggested in the literature—Carpenter et al.'s (1990) as complemented by Mackintosh and Bennett (2005), DeShon et al.'s (1995), and Dillon et al.'s (1981)—were tested in both studies using confirmatory factor analysis procedures. Analyses were also conducted in both studies to assess the extent to which gender differences supported the proposed taxonomies. It should be noted, however, that the student samples used are likely not representative of the general population. As such, this investigation does not address the issue of mean gender differences on the APM.

1. Study 1

1.1. Method

1.1.1. Subjects

The timed version of the APM was administered to 506 first-year students (326 women, 180 men) at the University of Toronto at Scarborough who were given extra credit in an introductory psychology course for their participation. Subjects ranged in age from 17 to 30 years (M=20.0, SD=1.8).

1.1.2. Procedure

In small-group settings, subjects completed both Set I and Set II of the APM. The standard instructions were read aloud by the experimenter. The standard timing of 5 min for Set I and 40 min for Set II were allowed. Only the results from Set II will be reported here.

1.2. Results

The descriptive statistics and the details from a principal component analysis of these 36-item APM Set II data can be found in Bors and Stokes (1998); more descriptive statistics for items (percent correct, skewnesses) were also reported by Vigneau and Bors (2005). It is worth recalling, however, that item skewnesses (M=−0.77, SD=1.86) ranged from −4.08 (item 2) to +4.61 (item 36) and that the test as a whole was internally reliable, α=0.84; Spearman–Brown corrected split-half correlation (odd–even)=0.72.

1.2.1. Factor analyses

To test the extent to which the inter-item covariance was accounted for by each of the three APM item taxonomies proposed in the literature, a number of measurement models were tested using confirmatory factor analysis. These models were a single-factor model, a correlated two-factor model based on Dillon et al.'s (1981) exploratory factor analysis, a correlated two-factor model based on DeShon et al.'s (1995) items classification, a correlated four-factor model based on Carpenter et al.'s classification as completed by Mackintosh and Bennett (2005), and a correlated two-factor model based on item skewnesses. The single-factor model and the skewness-based models were used as reference factor solutions. This last model was included in light of the role attributed to item position in the test in the interpretation of previous factor analytic results. In this two-factor model, all items whose skewnesses had negative values were assigned to a first factor (item 1 through item 23), whereas all items with positive skewnesses were assigned to a second factor (item 24 through item 36). All factor models were estimated using the maximum likelihood method as implemented in the SAS CALIS procedure (SAS, Version 9.1; SAS Institute, 2004). Covariance matrices were used as input data (an attempt was made at analyzing the tetrachoric correlation matrix, but this proved to be non-positive definite). All factor models were estimated using the maximum likelihood method. Eight goodness-of-fit indices were provided for each model. In addition to the χ^2 and its associated degrees of freedom, the goodness-of-fit index (GFI; Jöreskog & Sörbom, 1996), the comparative fit index (CFI; Bentler, 1990), the non-normed fit index (Bentler & Bonett, 1980), the standardized root mean square residual (SRMR), Akaike's information criterion (AIC; Akaike, 1987), Schwarz's Bayesian criterion (SBC, Schwarz, 1978) and the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993) are reported in Table 1 for each model.

Although most of the models' levels of fit cannot be directly compared statistically, inspection of the various fit indices of the five models tested globally revealed that their fits to the data were moderate and similar to one another. In fact, the only model emerging as achieving a relatively good fit was the skewness-based model. That is, a model based on statistical properties of the items and making no assumption relative to the type of information processing taking place during test completion, performed better as a measurement model than any of the models assuming specific information processing types for items.

1.2.2. Gender differences

On average, APM scores were higher for men (M=23.06, SD=5.67) than they were for women (M=21.68, SD =5.50), $t(504)=2.66$, $p<.05$, $d=0.24$. In correlational terms, APM scores were significantly associated with gender, $r=0.12$.

When 33 of the 36 APM items were grouped according to Mackintosh et al.'s (2005) classification (see Appendix A) to compute processing-type subscores, significant gender differences were observed for the quantitative-pairwise-progression composite, $t(504)=2.33$, $p<0.05$, and for the distribution-of-two-values composite, $t(504)=3.54$, $p<0.05$. Although being in the same direction, the gender differences did not reach significance for the figure-addition/subtraction and the distribution-of-three-values composites. These results are presented in terms of means, percent correct, standardized differences effect sizes (Cohen's d) and correlations in Table 2, along with the reliabilities of the composite scores.

Table 2
Male and female scores on the APM total score and eight APM item-type composite scores for men (n=180) and women (n=326)

	Men		Women		d	r
	Score	Percent correct	Score	Percent correct		
APM (36 items; alpha=0.84)	23.06 (5.67)	64.4	21.68 (5.50)	60.2	0.24	0.12*
Mackintosh and Bennett's (2005)						
PP (9 items; α=0.57)	6.88 (1.55)	76.4	6.52 (1.70)	72.4	0.20	0.10*
A/S (8 items; α=0.52)	5.86 (1.55)	73.3	5.79 (1.48)	72.4	0.04	0.02
D3 (9 items; α=0.60)	5.20 (1.81)	57.8	4.88 (1.84)	54.2	0.16	0.08
D2 (7 items; α=0.58)	2.53 (1.68)	36.1	2.01 (1.53)	28.7	0.32	0.16*
DeShon et al.'s (1995)						
Visual (13 items; α=0.68)	9.12 (2.32)	70.2	8.48 (2.33)	65.2	0.26	0.13*
Analytic (12 items; α=0.64)	6.48 (2.21)	54.0	6.12 (2.16)	51.0	0.16	0.08
Dillon et al.'s (1981)						
Pattern progression (17 items; α=0.67)	10.19 (2.68)	59.9	9.59 (2.57)	56.4	0.22	0.11*
Addition/subtraction (20 items; α=0.78)	13.26 (3.67)	66.3	12.36 (3.63)	61.8	0.25	0.12*
Skewness-based model						
Negatively skewed items (23 items; α=0.80)	18.93 (3.75)	82.3	18.19 (3.75)	79.1	0.18	0.09*
Positively skewed items (13 items; α=0.71)	4.13 (2.77)	31.8	3.49 (2.61)	26.8	0.23	0.11*

Note. * p < 0.05. Statistical tests' significance not adjusted for multiple testing. d = size of the gender effect expressed as Cohen's d; r = correlation between gender and score; PP = quantitative pairwise progression; A/S = figure addition/subtraction; D3 = distribution of three values; D2 = distribution of two values. Standard deviations in brackets.

Processing-type subscores based on DeShon et al.'s (1995) and Dillon et al.'s (1981) item classifications did not result in clearer, more differentiated, patterns of associations with gender. As could be predicted from theory, a visual processing score based on DeShon et al.'s (1995) classification was significantly related to gender, whereas an analytic processing score was not. However, both these correlations were low, and the difference between the two was not significant (z=0.452; correlation between the visual and analytic processing scores=0.67). Subscores based on Dillon et al.'s (1981) item classification were both significantly but weakly related to gender. Similar results were obtained with composite subscores from the skewness-based model (see Table 2).

The correlation between gender and score was computed for each of the 36 Set II item of the APM. These correlations are reported in Table 3, along with Mackintosh and Bennett's (2005) corresponding categories. As can be observed from the table, although the vast majority of the correlations were in the direction of a male advantage, they were all of low magnitude (smaller than 0.20). Five of these correlation were significant at the $p<0.05$ level. Interestingly, these top five correlations were associated, in terms of items categories, with a mix of distribution-of-two-values, quantitative-pairwise-progression, and distribution-of-three-values. Notably, no figure-addition/subtraction item was present even among the top 12 correlations with gender. This is surprising, given that this item category was reported by Mackintosh and Bennett (2005) to be one of the two with the largest gender difference. In fact, no obvious pattern of association between the gender effect and Mackintosh and Bennett's (2005) item categories emerged: Categories were scattered along the continuum of association between performance and gender. It could be argued that most of the distribution-of-two-values items appeared in the top half of the correlations (5 out of 7); but it would also have to be recognized that most figure-addition/subtraction were in the bottom half (5 out of 8), two of them even presenting values in the direction of a female advantage.

One consequence of these observations is that it is possible to construct subscores of APM items that will either correlate with gender, or not, and this irrespective of item category. For instance, a composite score comprised of distribution-of-two-values items 22, 23, 31, and 32 was relatively highly correlated with gender (r=0.19, $p<0.05$), whereas another distribution-of-two-values composite score, based on items 30, 32, 35, and 36 was not correlated with gender (r=0.06, p=0.19).

An analysis in terms of DeShon et al.'s (1995) processing types did not give any indication of a clearer pattern of item segregation: 6 out of the 13 items categorized as visual appeared in the bottom half of the rank-ordered correlations with gender. An analysis in terms of Dillon et al.'s (1981) two factors also provided mixed results, with both the addition/subtraction and the pattern-progression items being represented almost equally along the range of gender-score association. Thus, the analyses of this first data set did not provide much support for any of the information-processing based taxonomies of APM items.

Table 3
Association between gender and score for each APM (Set II) item, by decreasing magnitude of correlation (N=506)

Item no. (category)	r	Item no. (category)	r	Item no. (category)	r
22 (D2)	**.184**	9 (A/S)	.057	3 (PP)	.024
31 (D2)	**.130**	26 (PP)	.050	7 (A/S)	.024
18 (Unclas.)	**.106**	33 (A/S)	.047	13 (D3)	.024
5 (PP)	**.094**	28 (D3)	.045	12 (A/S)	.016
29 (D3)	**.093**	35 (D2)	.039	20 (A/S)	.013
27 (D3)	.082	19 (A/S)	.038	30 (D2)	.012
23 (D2)	.080	14 (PP)	.036	17 (D3)	.006
10 (PP)	.078	1 (D3)	.034	34 (D3)	.003
32 (D2)	.076	36 (D2)	.032	8 (D3)	−.007
21 (D3)	.068	24 (PP)	.029	2 (Unspec.)	−.033
6 (PP)	.063	11 (Unspec.)	.029	16 (A/S)	−.037
25 (PP)	.061	4 (PP)	.027	15 (A/S)	−.050

Note. Correlations in bold are significant ($p<0.05$). Statistical tests' significance not adjusted for multiple testing. PP = quantitative pairwise progression; A/S = figure addition/subtraction; D3 = distribution of three values; D2 = distribution of two values.

2. Study 2

2.1. Method

2.1.1. Subjects

A timed 14-item short version of the APM (VA-APM) was administered to 306 undergraduate university students (205 women, 101 men) at the University of Toronto at Scarborough. Subjects ranged from 17 to 54 years in age (M=21.1, SD=4.2). Fewer than 5% of the subjects were 27 or older. They were given extra credit in an introductory psychology course for their participation.

2.1.2. Materials

The 14-item short version (Vigneau & Bors, 2001) of the APM used in this study was designed based on DeShon et al.'s (1995) item taxonomy and results. It is comprised of seven items classified as visual (original APM item numbers 9, 12, 18, 22, 24, 32, and 33) and seven items classified as analytic (original APM item numbers 8, 13, 17, 21, 27, 30, and 34). These specific items were chosen as to represent a wide range of difficulty as well as similar difficulty levels across the two item categories. The items of the 14-item version were arranged in a test booklet in the order implied by their original item number (with one exception: item 9 was presented before item 8). Given its relation to DeShon et al.'s (1995) hypothesis of a distinction of APM items in terms of visual and analytic processes, this 14-item short version of the APM will be referred to as the verbal–analytic APM, or VA-APM, in the rest of the present article.

2.1.3. Procedure

The VA-APM was administered individually or in small-group settings of three or four individuals at a time. Standard APM instructions were used. Subjects completed two practice items from APM Set I before proceeding to the VA-APM. A time limit of 17 min to complete the 14-item VA-APM was used.

2.2. Results

The descriptive statistics for the 14 VA-APM items are presented in Table 4. The VA-APM total scores (M=7.62, SD = 2.90, SK = −0.10, KU = −0.28, minimum = 0, maximum = 14) appeared roughly normally distributed as well as moderately internally consistent, Cronbach's α = 0.70; Spearman–Brown corrected split-half reliability (odd–even) = 0.61.

2.2.1. Factor analyses

Four measurement models of the inter-correlations of the 14 VA-APM items were tested. These models were a single-factor model, a correlated two-factor model based on Dillon et al. (1981), a correlated two-factor model based on DeShon et al. (1995), and a correlated two-factor model based on item skewnesses. The four-factor model based on Mackintosh and Bennett (2005) could not be tested due to the limited number of items in some of the categories (in particular the quantitative-pairwise-progression category: only one item). As in Study 1, all factor models were estimated using the maximum likelihood method in SAS CALIS. In addition to the analyses of the item covariance matrix, the tetrachoric correlation matrix was also analysed. The results of these confirmatory factor analyses are presented in Table 5.

As was the case with the 36-item APM in Study 1, inspection of the selected fit indices revealed similar levels of fit for the four models tested. Once again, the only model emerging as achieving a somewhat better fit was the skewness-based model. Overall, model fit was higher in analyses of the covariance matrix than in analyses of the tetrachoric correlation matrix. However, the pattern of results was the same across both analysis strategies (see Table 5).

2.2.2. Gender differences

Although VA-APM total score was on average higher for men (M = 8.04, SD = 2.99) than it was for women (M = 7.42, SD = 2.84), the difference did not reach significance, $t(304)$ = 1.76, p = 0.08. The magnitude of the relation between total score and gender (r = 0.10, or d = 0.20) was similar, however, to that between the 36-item APM total score and gender in Study 1.

The VA-APM was originally designed to test assumptions of DeShon et al.'s (1995) classification, not Mackintosh et al.'s (2005). Although Mackintosh et al.'s (2005) model could not be tested with the present data, scores based on this model could nevertheless be computed for three of the four categories (as mentioned earlier, only one of the VA-APM items was classified as quantitative-pairwise-progression). Significant gender differences were found only for the figure-addition/subtraction composite score, with men (M = 2.06, SD = 0.88) obtaining higher scores than women (M = 1.73, SD = 0.88), $t(304)$ = 3.07, p < 0.05.

The 14 VA-APM items were classified according to the two categories of DeShon et al.'s (1995) taxonomy, to produce a visual processing score and an analytic processing score. Significant gender differences were found for the visual processing score, but not for the analytic processing score (see Table 6). The correlation between the latter and gender being virtually r = 0, it follows that the difference between the two score–gender correlations was also significant. Subscores based on Dillon et al.'s (1981) item classification also produced differential results, the addition/subtraction composite being significantly related to gender whereas the pattern-progression composite was not. Correlations between gender and composite scores from the skewness-based model were both in the direction of a male advantage, but did not reach significance (see Table 6).

These gender differences results at the composite scores level should be kept in perspective, especially in light of the fact that all reliability coefficients for composite scores derived from the VA-APM items were low. These low reliabilities, combined with the relatively high correlations

Table 4
Descriptive statistics for each of the 14 VA-APM items (N = 306)

Item no.	Original Item No.	M	SD	SK	KU
1	9	0.77	0.42	−1.32	0.26
2	8	0.89	0.32	−2.44	3.96
3	12	0.73	0.44	−1.05	−0.90
4	13	0.61	0.49	−0.44	−1.82
5	17	0.72	0.45	−0.98	−1.05
6	18	0.66	0.47	−0.68	−1.55
7	21	0.64	0.48	−0.59	−1.66
8	22	0.53	0.50	−0.12	−2.00
9	24	0.32	0.47	0.79	−1.38
10	27	0.38	0.49	0.49	−1.78
11	30	0.40	0.49	0.43	−1.83
12	32	0.35	0.48	0.65	−1.59
13	33	0.33	0.47	0.71	−1.51
14	34	0.30	0.46	0.87	−1.25

Table 5
Goodness-of-fit measures of four measurement models of the VA-APM (N = 306)

Model	χ^2 (df)	GFI	CFI	NNFI	SRMR	AIC	SBC	RMSEA (90% CI)
Single-factor model (14 items)								
1-factor	89 (77)	.96	.96	.96	.045	−65	−352	.023 (.000–.041)
1-factor	*333 (77)*	*.87*	*.75*	*.71*	*.073*	*179*	*−108*	*.104 (.093–.116)*
Dillon et al.'s (1981)								
2-factor (r = 0.84)	85 (76)	.96	.97	.97	.045	−67	−350	.020 (.000–.040)
2-factor (r = 0.94)	*331 (76)*	*.88*	*.75*	*.70*	*.072*	*179*	*−104*	*.105 (.094–.117)*
DeShon et al.'s (1995)								
2-factor (r = 0.94)	88 (76)	.96	.96	.96	.045	−64	−347	.023 (.000–.041)
2-factor (r = 0.96)	*332 (76)*	*.87*	*.75*	*.70*	*.072*	*180*	*−103*	*.105 (.094–.117)*
Skewness-based 2-factor model								
2-factor (r = 0.68)	67 (76)	.97	1.00	1.00	.039	−85	−368	.000 (.000–.024)
2-factor (r = 0.72)	*278 (76)*	*.90*	*.80*	*.77*	*.065*	*126*	*−157*	*.093 (.082–.105)*

Note. Results in roman are from analyses of the covariance (Pearson correlations) matrix; results in italics are form analyses of the tetrachoric correlation matrix. GFI = goodness-of-fit index; CFI = comparative fit index; NNFI = non-normed fit index; SRMR = standardized root mean square residual; AIC = Akaike's information criterion; SBC = Schwarz's Bayesian criterion; RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval for RMSEA. Dillon et al.'s model: one pattern-progression factor (6 items) and one figure-addition/subtraction factor (8 items). DeShon et al.'s model: one analytic factor (7 items) and one visual factor (7 items). Skewness-based model: one negative (items 1 through 8) and one positive (items 9 through 14) skewness factors. See Appendix for exact item assignations.

Table 6
VA-APM total score and seven VA-APM item-type composite scores for men (n = 101) and women (n = 205)

	Men		Women		d	r
	Score	Percent correct	Score	Percent correct		
VA-APM (14 items; α=0.70)	8.04 (2.99)	57.4	7.42 (2.84)	53.0	0.20	0.10
Mackintosh and Bennett's (2005)						
A/S (3 items; α=0.37)	2.06 (0.88)	68.7	1.73 (0.88)	57.7	0.34	0.17*
D3 (6 items; α=0.49)	3.54 (1.48)	59.0	3.53 (1.42)	58.8	0.00	0.00
D2 (3 items; α=0.39)	1.38 (1.03)	46.0	1.22 (0.96)	40.7	0.15	0.08
DeShon et al.'s (1995)						
Visual (7 items; α=0.55)	4.08 (1.68)	58.3	3.50 (1.66)	50.0	0.32	0.16*
Analytic (7 items; α=0.55)	3.96 (1.70)	56.6	3.92 (1.65)	56.0	0.02	0.01
Dillon et al.'s (1981)						
Pattern progression (6 items; α=0.44)	2.75 (1.24)	55.0	2.70 (1.19)	54.0	0.04	0.02
Addition/subtraction (8 items; α=0.64)	5.29 (2.09)	58.8	4.72 (2.12)	52.4	0.25	0.13*
Skewness-based model						
Negatively skewed items (8 items; α=0.63)	5.78 (1.92)	72.3	5.43 (1.89)	67.9	0.18	0.09
Positively skewed items (6 items; α=0.52)	2.26 (1.62)	37.7	1.99 (1.50)	33.2	0.16	0.08

Note. $^*p<0.05$. Statistical tests' significance not adjusted for multiple testing. d=size of the gender effect expressed as Cohen's d; r=correlation between gender and score; A/S = figure addition/subtraction; D3 = distribution of three values; D2 = distribution of two values. Standard deviations in brackets.

between composite scores (between the visual and analytic composites based on DeShon et al. (1995): $r=0.50$; between the pattern-progression and addition/subtraction composites based on Dillon et al. (1981): $r=0.48$) mean that the differential predictive power of any of the subscores considered here, relative to gender, was bound to be low. Also, all correlations with gender were weak ($r\leq0.17$, or $d\leq0.34$). In fact, none of the composite scores reached correlations with gender higher than the highest-correlating single item (VA-APM item 13, original APM item number 33; see Table 7).

3. General discussion

Both studies reported in the present article were aimed at assessing the validity of three information-processing-based taxonomies of APM items, namely Carpenter et al.'s (1990) taxonomy as completed by Mackintosh and Bennett (2005), DeShon et al.'s (1995), and Dillon et al.'s (1981). In each of the two studies, confirmatory factor analyses and gender differences analyses were conducted. Despite differences in versions of APM used—the 36-item Set II of the APM in Study 1 vs. a 14-item version based on DeShon et al. (1995) in Study 2—the results of the two studies were, in the main, consistent and illustrated the limitations of all three taxonomies examined.

The confirmatory factor analyses in both studies failed to provide substantial support for any of the three APM item taxonomies examined. In some instances, the fit of the more complex models was marginally better than the fit of a simple one-factor model, but in all cases the correlations between the conceptual factors were high enough to question the heuristic value of the distinctions being made. In fact, the best fit overall was obtained, in both studies, not by a cognitive theory-based model, but by a two-factor model based on the skewness of the item distributions.

The fact that a skewness-based two-factor measurement model was the best-fitting model is not straightforwardly interpreted. As indicated in the introduction, item skewness in the APM is confounded with both item position in the test and item difficulty. In turn, several psychological factors may be associated with item position to influence performance, including learning effects and fatigue. In timed administrations of the APM, item difficulty may be the product of time pressure and differential response rates as much as of information-processing properties of the items. In the absence of experimental manipulations of the test, such as administering APM items in various orders, or removing the time limit, the various potential sources of item segregation in terms of skewness are impossible to disentangle.

It is sometimes argued that factors based on skewness could result from using inappropriate statistical procedures when analyzing dichotomous data. Specifically, the standard linear model, as it is used for instance in principal component analysis or confirmatory factor analysis, requires, among other things, that variables be normally distributed. This assumption is clearly violated in the case of the dichotomously-scored and sometimes substantially skewed APM items. It is commonly assumed that, although parameter estimates derived with estimation methods that assume normality, including maximum likelihood, are fairly accurate when data are non-normal, standard errors and test statistics may be incorrect (Bollen, 1989). More specifically, significance tests tend to be significant too often (Kline, 1998), that is, the correct model is rejected too often. However valid such conclusions drawn mainly from simulation studies may be, potential significance test bias associated with variable non-normality probably has a limited impact on the conclusions of the present investigation of the dimensionality of the APM. Indications that APM factors based on item position in the test are more than just statistical artifacts associated with linear

Table 7
Association between gender and score for each VA-APM item, by decreasing magnitude of correlation ($N=306$)

33 (A/S)	12 (A/S)	18 (U)	21 (D3)	32 (D2)	13 (D3)	9 (A/S)	22 (D2)	24 (PP)	30 (D2)	34 (D3)	8 (D3)	27 (D3)	17 (D3)
.18	.10	.08	.08	.07	.07	.06	.05	.05	.03	.03	.01	−.08	−.09

Note. Correlation in bold is significant ($p<0.05$). Statistical tests' significance not adjusted for multiple testing. Original APM item numbers. PP = quantitative pairwise progression; A/S = figure addition/subtraction; D3 = distribution of three values; D2 = distribution of two values.

analysis of dichotomous data have been provided by Vigneau and Bors (2005). In an item response theory analysis of the present Study 1 data, they found that although the results supported a multidimensional model of the APM, they were also consistent with multiple unidimensional item clusters based on item position. Exploratory factor analysis (Bors & Stokes) as well as non-linear factor analysis (Vigneau and Bors, 2005) of the same data also corroborated this interpretation. In any case, it seems that until more experimental approaches are used in combination with factor analyses of item responses, the dimensionality of the APM will remain obscure.

With respect to the Sample 2 data, in which the tetrachoric correlation matrix could be analysed, results also indicated that the hierarchy of model fit was consistent across analysis methods. That is, although the differences in levels of fit between the covariance and the tetrachoric correlation analyses were substantial in absolute terms (see Table 5), both strategies lead to the same conclusions in relative terms. That is, regardless of the analytic strategy used, Dillon et al.'s (1981) and DeShon et al.'s (1995) models exhibited levels of fit virtually identical to one another and to a simple one-factor model, and the skewness-based model achieved a significantly better fit than all the other models examined.

Analyses of the proposed APM item taxonomies in terms of gender differences did not yield much evidence supporting either the value of the proposed taxonomies or the relevance of gender differences for the uncovering of cognitive processes involved in performance on the APM. We have not been able to replicate Mackintosh and Bennett's (2005) reanalysis of their own data. According to Mackintosh and Bennett (2005) we should have observed a pronounced male advantage for figure-addition/subtraction and distribution-of-two-values composite scores. Instead, we saw in Study 1 only a small male advantage on the distribution-of-two-values and on the quantitative-pairwise-progression composites, and in Study 2, in which no quantitative-pairwise-progression composite could be computed, we only observed a small male advantage on the figure-addition/subtraction composite. These inconsistent results provide very limited support for Mackintosh and Bennett's (2005) hypotheses. Although limited statistical power could be invoked to explain the failure to reveal a gender effect in some item clusters, the inconsistencies in the gender effects actually identified are entirely consistent with the recent analyses of data from 1970 young adults reported by Colom and Abad (2007), who observed a small and non-specific gender effect for all four item categories based on Mackintosh and Bennett (2005). In this context, the low reliabilities of the composite scores based on Mackintosh and Bennett (2005) must be emphasized, especially with respect to our Study 2 data, in which the values of the alpha reliabilities were all below 0.5.

Gender differences analyses in terms of DeShon et al.'s (1995) and Dillon et al.'s (1981) taxonomies did not provide stronger support for their proposed item classifications. In both cases, weak associations with gender were observed, and results were again plagued by unacceptably low reliabilities of the composite scores. A stronger case was found for the composite scores based on DeShon et al. (1995). Here, consistent results of an association between the visual composite and gender in the absence of an association between the analytic composite and gender replicated across the two studies. However, the value of these results remain questionable in light of the magnitude of the obtained gender effect (r=0.17, or d=0.34, that is, less than 3% of shared variance), the low reliabilities of the composites, and the substantial correlation between the composites. Such conditions indeed leave little room for any predicting specificity of the proposed processing scores.

Finally, item analyses also underscored the limitations of item processing categories by revealing their internal heterogeneity. The fact that the APM items did not cluster according to processing categories in their association with gender was an additional indication of the limitations of the proposed taxonomies. This inconsistency, however, was in line with our factor analytic results as well as with the observed low reliabilities of the processing composite scores.

In conclusion, the identification of APM performance dimensions based on information processing continues to be an open question and is likely to remain intractable without experimental manipulations of important variables related to the construction of the items and to the test as a whole. Additionally, given that items may be solved using different strategies, it would be helpful to begin by identifying such strategies and any cross-item consistencies. Furthermore, observed gender differences appear to be distributed across all items on the APM, rather than forming clusters. Although a precise estimation of the strength of these gender differences in the population was not the main objective of the present research and could only be achieved through the use of population-representative samples, it seems to vary from study to study. Given present and past findings, the distribution of the male advantage across items may very well be found to be different yet again in the next data set. Attempts to identify factors on the APM by associating gender differences with performance on particular items and clusters of items do not appear to be a fruitful venture.

Appendix A

Four categorizations of Raven's APM Set II items

Item no.	Carpenter[a]	DeShon[b]	Mackintosh[c]	Dillon[d]
1	D3	Analytic	[D3]	[PP]
2	Not included	Either	Unspecified	Pattern progression
3	PP	Visual	PP	Pattern progression
4	PP	Analytic	PP	Pattern progression
5	PP	Either	PP	Pattern progression
6	PP	Either	PP	[A/S]
7	A/S (addition)	Visual	A/S	Addition/subtraction
8	D3	Analytic	D3	[PP]
9	A/S (addition)	Visual	A/S	Addition/subtraction
10	PP	Visual	PP	Addition/subtraction
11	Not included	Visual	Unspecified	Addition/subtraction
12	A/S (subtraction)	Visual	[A/S]	[A/S]
13	D3	Analytic	D3	[A/S]
14	PP	Either	PP	[PP]
15	Not included	Uncodable	A/S	[A/S]
16	A/S (subtraction)	Visual	A/S	Addition/subtraction
17	D3	Analytic	D3	Pattern progression
18	Unclassified	Visual	Unclassified	[A/S]
19	Unclassified	Both	A/S (Addition)	[A/S]
20	Not included	Both	A/S	[PP]

(continued on next page)

Appendix A (*continued*)

Item no.	Carpenter[a]	DeShon[b]	Mackintosh[c]	Dillon[d]
21	Not included	Analytic	D3	Addition/subtraction
22	D2	Visual	D2	[A/S]
23	D2	Visual	D2	[A/S]
24	Not included	Visual	PP	[A/S]
25	Not included	Both	PP	[A/S]
26	PP; D3	Both	PP	Pattern progression
27	D3	Analytic	D3	[PP]
28	Not included	Analytic	D3	Addition/subtraction
29	D3	Analytic	D3	[PP]
30	Not included	Analytic	D2	[PP]
31	D3; D2	Both	D2	[A/S=PP]
32	D3; D2	Visual	D2	[PP]
33	A/S	Visual	A/S	[PP]
34	D3	Analytic	D3	[A/S]
35	D2	Both	D2	Addition/subtraction
36	D2	Analytic	[D2]	Pattern Progression

[a]Carpenter et al.'s (1990) classification. “Constant in a row” category omitted.

[b]DeShon et al.'s (1995) processing-type classification.

[c]Mackintosh and Bennett's (2005) classification, following their Appendix A and exceptions indicated in their Table 1. Square brackets indicate items not mentioned explicitly by Mackintosh and Bennett's (2005) and thus categorized according to the statement that they classified items following the Carpenter et al. (1990) rules.

[d]Dillon et al.'s (1981) two subsets of items recommended as relatively pure indicators of a pattern addition/subtraction factor and a pattern progression factor. Square brackets indicate the factor with the highest loading (irrespective of magnitude) for the remaining items in the factor solution reported by Dillon et al. (1981).

References

Abad, F. J., Colom, R., Rebollo, I., & Escorial, S. (2004). Sex differential item functioning in the Raven's Advanced Progressive Matrices: Evidence for bias. *Personality and Individual Differences*, *36*, 1459–1470.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

Alderton, D. L., & Larson, G. E. (1990). Dimensionality of Raven's Advanced Progressive Matrices items. *Educational and Psychological Measurement*, *50*, 887–900.

Arthur, W., & Woehr, D. J. (1993). A confirmatory factor analytic study examining the dimensionality of the Raven's Advanced Progressive Matrices. *Educational and Psychological Measurement*, *53*, 471–478.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, *58*, 382–398.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. J. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). London: Sage.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, *97*, 404–431.

Colom, R., & Abad, F. J. (2007). Advanced progressive matrices and sex differences: Comment to Mackintosh and Bennett (2005). *Intelligence*, *35*, 183–185.

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence*, *21*, 135–155.

Dillon, R. F., Pohlmann, J. T., & Lohman, D. F. (1981). A factor analysis of Raven's Advanced Progressive Matrices freed of difficulty factors. *Educational and Psychological Measurement*, *41*, 1295–1302.

Hunt, E. B. (1974). Quote the Raven? Nevermore! In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 129–158). Hillsdale, NJ: Erlbaum.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's Reference Guide.* Chicago: Scientific Software International.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Kubinger, K. D., Formann, A. K., & Farkas, M. G. (1991). Psychometric shortcomings of Raven's Standard Progressive Matrices, in particular for computerized testing. *Revue européenne de psychologie appliquée*, *41*, 295–300.

Lynn, R., Allik, J., & Irwing, P. (2004). Sex differences on three factors identified in Raven's Standard Progressive Matrices. *Intelligence*, *32*, 411–424.

Mackintosh, N. J., & Bennett, E. S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, *33*, 663–674.

Raven, J. C. (1956). *Progressive Matrices, Sets A, B, C, D and E.* London: H. K. Lewis.

Raven, J. C. (1962). *Advanced Progressive Matrices, Set II.* London: H. K. Lewis.

SAS Institute (2004). *SAS/STAT User's Guide (Version 9.1) [software manual].* Cary, NC: Author.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental Rotation, a group test of three-dimensional spatial visualization. *Perceptual & Motor Skills*, *47*, 599–604.

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.

Vigneau, F., & Bors, D. A. (2001). Visuospatial and verbal–analytic processes in Raven's Progressive Matrices performance. *XIth Meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science (CSBBCS), Université Laval, Quebec City.*

Vigneau, F., & Bors, D. A. (2005). Items in context: Assessing the dimensionality of Raven's Advanced Progressive Matrices. *Educational and Psychological Measurement*, *65*, 109–123.

Voyer, D., Voyer, S. D., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.