Available online at www.sciencedirect.com

SCIENCE DIRECT®

Intelligence 33 (2005) 663–674

INTELLIGENCE

What do Raven's Matrices measure? An analysis in terms of sex differences

N.J. Mackintosh*, E.S. Bennett

University of Cambridge, Department of Experimental Psychology, Downing Street, Cambridge CB2 3EB, England, United Kingdom

Received 1 July 2004; received in revised form 2 January 2005; accepted 2 March 2005
Available online 23 May 2005

Abstract

Although it is sometimes claimed that Raven's Matrices provide an almost pure measure of *g*, there is evidence that the easier items in the Standard Progressive Matrices and in Set I of the Advanced Matrices measure a perceptual or Gestalt factor distinct from the more analytic items in the rest of the tests. There is also, however, both factor analytic and experimental evidence that these analytic items fall into two partially separate groups, distinguishable by the type of rule needed for their solution in the analysis proposed by Carpenter, Just, and Schell (1990) [Carpenter, P. A., Just, M. A., Schell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven's Progressive Matrices Test. *Psychological Review*, *97*, 404–431]. Re-analysis of published data suggests a new source of evidence for this further distinction: males do better than females on items requiring an addition/subtraction or distribution of two rule, but there is no sex difference on items requiring pairwise progression or distribution of three rules. A specially designed experiment confirmed this pattern of results.

Keywords: Intelligence; Raven's Matrices; Sex differences

1. Introduction

Raven's Matrices were initially developed as a measure of the eduction of relations and correlates posited by Spearman (1927) to underlie general intelligence or *g* (Penrose & Raven, 1936; Raven, 1938).

* Corresponding author.
E-mail address: n.mackintosh@psychol.cam.ac.uk (N.J. Mackintosh).

0160-2896/$ - see front matter
doi:10.1016/j.intell.2005.03.004

The original Standard Progressive Matrices (hereafter abbreviated to RPM) were later supplemented by the easier Coloured Progressive Matrices (RCPM) and the more difficult Advanced Progressive Matrices (RAPM). Spearman continued to regard Raven's as "perhaps the best of all non-verbal tests of *g*" (Spearman, 1946, p. 127), a view endorsed by other British psychometricians such as Vernon, who regarded Raven's as "an almost pure *g* test" (Vernon & Parry, 1949, p. 234). This strong claim was later endorsed by Jensen (1980), who wrote that "factorially the [R]PM apparently measures *g* and little else... The loadings that are occasionally found on other 'perceptual' and 'performance'-type factors, independently of *g*, are usually so trivial and inconsistent from one analysis to another as to suggest that the RPM does not reliably measure anything but *g* in the general population" (p. 646). The only authority cited by Jensen as support for this claim was a review by Burke (1958), which is surprising since Burke had argued that this was a peculiarly English view sponsored by Spearman, and had concluded his review of factorial studies with the statement: "The evidence is not convincing that [R]PM (1938) has validity as a pure measure of the Spearman construct of *g*" (p.212).

Jensen has continued to argue that probably the "most distinctive feature" of Raven's test "is its very low loadings on any factor other than *g*" (Jensen, 1998, p. 38). But there now seems to be good evidence, not only from factor analytic studies, that the various forms of the Matrices tests measure at least two distinguishable processes, one that has been variously called figural, perceptual or Gestalt, the other usually termed analytic or analogical. We shall refer to them as perceptual and analytic. A Rasch analysis of Sets A to E of RPM concluded that Set A and the first half of Set B measured the perceptual process, while the second half of Set B and Sets C to E measured the analytic process—although other processes were involved in the solution of half the items in Set E (Van der Ven & Ellis, 2000). More recently, Lynn, Allik, and Irwing (2004) factor analysed the RPM scores obtained by some 2700 Estonian 12- to 18-year-olds, and confirmed van der Ven and Ellis's identification of a perceptual factor measured by Set A and early items in Set B. However, they also found that van der Ven's analytic factor could be separated into two distinct factors. We return to this latter distinction below.

In an analysis of the 12 items in Set I of RAPM, Hunt (1974) argued that a perceptual algorithm was sufficient to solve items 1–6, while an analytic algorithm was needed to solve items 7–12 (and would also solve items 1–6). Neuropsychological evidence further supports the distinction between perceptual and analytic processes. Villardita (1985) tested patients with right or left hemisphere damage on RPM and RCPM. Patients with right hemisphere damage performed relatively poor on perceptual items, while those with left hemisphere damage performed relatively poor on analytic items. In a complementary fMRI study, Prabhakaram, Smith, Desmond, Glover, and Gabrieli (1997) found that easier perceptual items predominantly activated the right frontal cortex, while harder analytic items resulted in bilateral frontal activation.

Is this distinction between relatively easier perceptual items and relatively harder analytic items the only difference between different Raven's items? Set II of RAPM does not contain easy figural items, and some confirmatory factor analyses of RAPM have concluded that a two-factor solution does not provide a better fit than a one-factor solution (Abad, Colom, Rebollo, & Escorial, 2004; Arthur & Woehr, 1993). However, Dillon, Pohlmann, and Lohman (1981) reported a two-factor solution, and, as noted above, in their factor analysis of RPM, Lynn et al. (2004) concluded that Van der Ven and Ellis (2000) analytic factor should be divided into two: the later items in Set B, and all but the last items in Sets C and D loaded onto one factor, while these last items in Sets C and D and nearly all items in Set E loaded onto a second. And in an experimental study of college students' performance on RAPM, DeSchon, Chan, and Weissben (1995) distinguished between two classes of items—those that required a visuo-spatial rule for their solution and those that required a verbal–analytic rule. The former, but not the latter, showed a "verbal overshadowing" effect: performance on the visuo-spatial items suffered when participants were required to describe out loud the nature of each item and how they were setting about solving it; performance on the verbal–analytic items did not.

We believe that the nature of the distinction between these two analytic factors or types of rule may be best understood in terms of the rules inferred by Carpenter, Just, and Schell (1990) as being needed to solve RAPM (indeed, DeShon et al.'s analysis follows Carpenter et al.'s quite closely). In a computational and experimental study, Carpenter et al. concluded that at least five distinct rules were used to solve the 36 items in Set II of RAPM (with a small number of items remaining "unclassified"). The rules were as follows:

1. Constant in a row: the same value of an attribute appears in all three figures in a row, but changes between rows.
2. Distribution of three: a different value of an attribute appears in each of the three figures of a row.
3. Pairwise progression: a constant change occurs in the size, number or position of an attribute between neighbouring figures in a row.
4. Addition or subtraction: a figure in one column is added to, or subtracted from, a figure in a second to produce the third.
5. Distribution of two: two values of an attribute occur in each row, with the third value being null.

As Carpenter et al. (1990) noted, the distinction between these rules is not perfectly clear cut: for example, pairwise progression in effect involves three different values of an attribute in each row—i.e., is reducible to rule 2; while addition/subtraction will also solve distribution of two problems. Be this as it may, the distinction between rules 1, 2 and 3 on the one hand and rules 4 and 5 on the other, does map quite closely onto the distinction between the verbal–analytic and visuo-spatial items identified by DeShon et al. (1995)—and also onto the two analytic factors identified by Lynn et al., 2004 in their factor analysis of RPM.[1] Rules 1, 2 and 3 in Carpenter et al.'s (1990) analysis, constant in a row, distribution of three and pairwise progression, are sufficient to solve the items in Sets B, C and D of RPM that load onto one of Lynn, Allik, and Irwing's factors; while, although the fit is not perfect, most of the items that load onto their second analytic factor (i.e., Set E and the last items in Sets C and D) require the application of rules 4 or 5, addition/subtraction or distribution of two. Similarly, 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 visuo-spatial. The one discrepancy is that DeShon et al. classify some rule 3 (pairwise progression) items as visuo-spatial and others as verbal–analytic.

Evidence from sex differences supports the reality both of the original distinction between perceptual and analytic factors, and of that between the two analytic factors (regardless of the labels attached to them). Lynn et al. (2004) reported that, at age 17, when there was an overall male superiority in total RPM score (see Lynn, Allik, Pullman, & Laidra, 2002), males obtained higher scores than females on their perceptual factor and on one of their two analytic factors, but not on the

[1] Confusingly enough, however, the *labels* Lynn et al. (2004) give to these two factors are diametrically opposed: the factor they label as a verbal–analytic reasoning corresponds most closely to the kind of item De Schon et al. (1995) describe as visuo-spatial; while the factor they label as visuo-spatial ability corresponds most closely to the items identified by DeSchon et al. as verbal–analytic.

other. If there is a sex difference on one analytic factor, but not on another, the factors must be measuring somewhat different processes. In a study of white South African university students, Rushton and Skuy (2000) found a small but statistically significant male advantage in total RPM score (90.9% correct vs. 89.2% for females). However, further analysis of their data, kindly made available by Rushton, reveals that this sex difference was confined to a small number of items, namely those that load onto the second of the two analytic factors in Lynn, Allik, and Irwing's analysis. Their confirmatory factor analysis showed that items C 12, D 12, and E 7–12 had high loadings on this factor and negligible loadings on either the perceptual or the other analytic factor: on these eight items, males obtained a score of 64.4% and females 53.1% in the Rushton and Skuy study. On the remaining 52 items in the test, male and female scores were 94.7% and 94.9%, respectively. There is an obvious problem of a ceiling effect here—which is particularly true for the perceptual items in Sets A and B, where males and females both averaged over 97% correct. This will inevitably have obscured any sex difference on the perceptual factor. Nevertheless, there is clear evidence that males did do better than females on those items that engaged the addition/subtraction or distribution of two rules in Carpenter et al. (1990) analysis, just as the 17-year-olds did in the Lynn, Allik, and Irwing study.

We have also analysed some data collected by ourselves (Mackintosh & Bennett, 2003), where this confound with item difficulty is less apparent. We administered RAPM to 34 male and 95 female English sixth-form students, aged 16–17. Item 1 was given as a practice item, and the remaining 35 items were given with a 30-min time limit. The results are shown in Table 1, with the items classified by the Carpenter et al. (1990) rules. We have omitted the constant in a row classification, since all items in Set II of RAPM also require at least one further rule for their solution. We have also, as noted in the table, reclassified some of the items left unclassified by Carpenter et al. (1990). Overall, males did rather better than females, but they did less well on distribution of three items (and on the one remaining unclassified item), slightly better on pairwise progression items, and obtained substantially higher scores on only two categories: addition/subtraction and distribution of two. These results seem to confirm the suggestion derived from the RPM data, that male advantage on Raven's is largely confined to those items requiring one or the other of these two rules. In this case, there can be no question that male advantage will be found only on the harder items: the addition/subtraction items, which yielded male superiority, were substantially easier than those requiring the distribution of three

Table 1
Male and female scores (percent correct) on different types of RAPM items (data from Mackintosh & Bennett, 2003)

Test	Males	Females
Total	60.4	57.0
D3	40.5	42.7
PP	76.3	72.8
A/S	76.9	67.3
D2	30.6	21.9
U	58.8	66.3

Total—total Raven's score; PP—pairwise progression; D3—distribution of three; A/S—addition/subtraction; D2—distribution of two; U—unclassified.

The item types are based on the analysis of Carpenter et al. (1990), with the following exceptions: we have classified Item 19 as figure addition, Item 21 as distribution of three, Item 25 as pairwise progression, and Item 30 as distribution of two. They were all unclassified in the original analysis. This leaves only Item 18 unclassified.

rule, where females outscored males. There is, however, another potential confound—item order. Although it is not true for the addition/subtraction items, the distribution of two items, which also show a male advantage, tend to come toward the end of the test. Since there was some evidence that males were responding more rapidly than females (for example, 71% of males, but only 65% of females, attempted Item 36, the last in the test), it is conceivable that males did better on the distribution of two items because they had more time for them.

The present experiment was designed as an explicit test of the proposition that males obtain higher scores than females only on certain Raven's items, in particular those that require for their solution the addition/subtraction or distribution of two rules in Carpenter et al. (1990) analysis. We are inclined to doubt the reliability of the small difference seen in Table 1 on pairwise progression items. We constructed two special sets of 18 items drawn from Set E of RPM and Set II of RAPM. After a few practice items, participants were allowed 20 min to complete each set (this was sufficient for essentially everyone to complete each set). As a further precaution, we departed from the standard Raven's procedure of progressing from easier to harder items: the relatively difficult items of special interest were placed in the middle of each set, which began and ended with easier items. Participants also completed two other tests, the DAT Verbal Reasoning test, and Vandenberg and Kuse's Mental Rotation test.

2. Method

The participants were 97 students from two sixth-form colleges in Cambridge (which take in students in the last 2 years of secondary education). There were 48 males and 49 females aged 17/18 years, all of whom had competed the first year of the sixth form (they were thus the equivalent of American high school seniors). They were tested in class sizes of 6 to 18 in two separate sessions, each of approximately 45 min duration. The following three tests were completed by all participants.

2.1. Raven's Matrices

The study required approximately equal numbers of Raven's items from the four Carpenter et al. (1990) categories listed in Table 1. In order to achieve this, we substituted items E7, E8, E10, E11 and E12 from RPM (all either addition/subtraction or distribution of two items) for items 1, 2, 11, 12, and 36 from Set II of RAPM, the first four of which are pairwise progression or distribution of three items. Item 36 is distribution of two, but was omitted because, in our experience, it is solved by very few participants. The resulting 36 items were divided into two sets of 18, which were ordered so that the most difficult items and the different Carpenter et al. categories were distributed more or less equally throughout each test. Full details of the two sets are given in Appendix 1.

A practice set of four questions, one from RAPM, three from RPM was given. These included two pairwise progression and two distribution of three items. Question 1 was worked through with the participants who were then instructed to solve the remaining three as quickly as possible. Answers were checked.

It was important that all items should be completed by all participants. They were instructed to do so, even if their answer was a best guess. On the basis of some pilot work, we allowed 20 min for each set. A break of 2 min was allowed between the two sets.

2.2. *The differential aptitude test (DAT) of verbal reasoning (1990)*

Participants were given 20 min to complete 40 questions.

2.3. *MRT-A from Vandenberg and Kuse (1978) mental rotations test*

This was the standard Set A, which contains 24 problems. Each problem shows a three-dimensional target figure paired with four choice figures, two of which are rotated versions of the target figure. To score a point, both correct answers must be given. The first 12 problems were attempted in 4 min with a 2-min break before attempting the second 12 in another 4 min. (The test is normally completed by allowing 3 min for each part. However, pilot studies indicated that this was too short a time for 17/18-year-olds.)

For all participants, testing was carried out over 2 sessions. They completed the MRT and DAT tests in the first session and, 1 week later, the two Raven's sets in the second.

3. Results

Table 2 shows the means and standard deviations of the number of correct responses given by male and female students on Raven's (both sets combined), MRT and DAT verbal reasoning tests. The final two columns give the effect size of the difference between male and female scores, and a *t*-test of the significance of the difference. In the most recent standardization of the DAT in Britain (Hyde & Trickey, 1995), 18-year-old male students obtained a score of 25.04 and females one of 24.87. Our own data show an equally small sex difference—and only slightly higher mean scores. The difference between males and females on the Vandenberg and Kuse MRT is highly significant and of approximately the same size as that seen in other samples (Masters & Sanders, 1993). The difference between males and females on our Raven's test was also substantial, equivalent to 6–7 IQ points, and thus rather larger than that seen in other studies. We comment on this below.

Table 3 shows the scores obtained by males and females on the four types of Raven's item. Males did better than females on all four types, with the largest numerical difference being on distribution of two items, closely followed by the difference on addition–subtraction items. Because of the greater variance on the distribution of two items, the effect size for addition–subtraction items was in fact larger, but both

Table 2
Male and female scores on Raven's, MRT and DAT tests

Test	Male (N=48)		Female (N=49)		Effect size	*t*-test
	Mean	S.D.	Mean	S.D.		
Raven's (36)	25.21	5.97	22.94	4.31	0.43	2.15*
MRT (24)	16.48	4.57	12.24	4.62	0.84	4.54**
DAT (40)	27.90	7.14	27.16	7.71	0.10	0.48

The number in brackets refers to the number of items in the test.
* $p<0.05$.
** $p<0.01$.

Table 3
Male and female scores on four types of Raven's items

Item type	Male		Female		Effect size	*t*-test
	Mean	S.D.	Mean	S.D.		
Pairwise progression (9)	7.35	1.34	7.12	1.28	0.17	0.87
Distribution of three (8)	4.92	1.65	4.78	1.03	0.10	0.51
Addition/Subtraction (10)	7.60	1.77	6.61	1.34	0.60	3.16**
Distribution of two (8)	4.58	2.41	3.55	2.05	0.45	2.24*

* $p<0.05$.
** $p<0.01$.

were significant. By contrast, the differences between males and females on distribution of three and pairwise progression items were small and not significant.

The finding that two differences were significant and two were not is not, of course, sufficient to establish that the two former differences were significantly larger than the latter two. That requires the demonstration of a significant interaction between sex and type of item. Because there were slightly different numbers of each type of item, we converted the participants' scores to percent correct, and performed an overall analysis of variance, which revealed a significant sex difference, $F(1, 95)=5.12$, $p<0.05$, a significant difference in performance on the four types of item, $F(3, 285)=73.03$, $p<0.001$, and a significant interaction between these two factors, $F(3, 285)=3.24$, $p<0.05$.

It is important to note that this pattern of sex differences was not confounded by differences in overall level of difficulty. It is true that the largest numerical difference between male and female scores was on the category with the lowest level of performance, distribution of two, and that the easiest type of item, pairwise progression, yielded no sex difference. But there was no difference on the distribution of three items, which were substantially harder than addition–subtraction items which yielded the largest effect size.

We also calculated the correlations, for males and females separately, between MRT and DAT scores on the one hand, and total Raven's scores as well as scores on the different categories of Raven's item. These are shown in Table 4. The most notable feature of the overall correlations between the three tests is that males' Raven's performance correlated more highly with their MRT than with their DAT scores, while the opposite was true for females (for both males and females, there was only a modest correlation between

Table 4
Correlations between MRT, DAT and Raven's scores for males and females

	Total	PP	D3	A/S	D2
Males					
MRT	0.62**	0.42**	0.64**	0.40**	0.52**
DAT	0.35*	0.13	0.33*	0.26	0.39**
Females					
MRT	0.40**	0.35*	0.21	0.25	0.29*
DAT	0.50**	0.49**	0.16	0.41**	0.33*

Total—total Raven's score; PP—pairwise progression; D3—distribution of three; A/S—addition/subtraction; D2—distribution of two.
* $p<0.05$.
** $p<0.01$.

MRT and DAT scores, for males, $r=0.32$, $p<0.05$; for females, $r=0.17$, ns). Table 4 also shows that what was true for total Raven's scores was, by and large, also true for scores on the different categories of Raven's items. In males, the correlations between MRT and all four types of item were highly significant, and larger than any correlation with DAT scores. In females, three of the four types of item correlated more strongly with DAT than with MRT, and in the fourth case neither correlation was significant. It should be acknowledged, however, that (perhaps partly owing to our relatively small sample size) none of the differences between these correlations described in this paragraph were statistically significant.

4. Discussion

That there was no significant difference between male and female scores on the DAT verbal reasoning test is consistent with the results of the latest British standardization of the DAT (Feingold, 1988; Hyde & Trickey, 1995, also reported no difference between American 18-year-olds in the 1980 U.S. standardization; but see Colom & Lynn, 2004, for contrary results from the recent Spanish standardization). The significant male advantage we observed on the MRT is also consistent with numerous other studies (Masters & Sanders, 1993). The substantial difference we found in overall Raven's scores, equivalent to an IQ difference of 6–7 points, is perhaps more surprising. It is certainly larger than any we have previously observed with this subject population (see Table 1 for data from Mackintosh & Bennett, 2003; in other, unpublished studies we have found even smaller differences and in one a small female advantage). It is also larger than the largest difference reported in any other study of Raven's performance in this general age group. Given their similar scores on the DAT, it is unlikely that the present sample was inadvertently biased in favor of high-scoring males. The most plausible explanation is that the test we constructed from Raven's items contained an unusually high proportion of those items (addition/subtraction and distribution of two) that favor males. A small male advantage on Raven's is not surprising. In an early review of the literature, Court (1983) concluded that studies of RPM and RAPM had yielded all possible outcomes—male advantage, female advantage, and no difference. In fact, however, studies of older participants (over the age of 16) were more likely to yield male than female advantage, and some, but by no means all, later studies have also found a slight male advantage (Colom & Garcia-Lopez, 2002; Flynn, 1998; Lynn et al., 2002; Rushton & Skuy, 2000; see Lynn, 1998; Mackintosh, 1998, for a discussion of the evidence).

Much the most important result of this experiment, however, is that males obtained significantly higher scores than females only on certain types of Raven's items, namely, those that require for their solution the addition/subtraction or distribution of two rules identified by Carpenter et al. (1990). On other kinds of item, there was no sex difference in performance. Although our sample size was not large, our confidence in our results seems justified since they are entirely consistent with those obtained in earlier studies of RPM and RAPM (Lynn et al., 2004; Mackintosh & Bennett, 2003; Rushton & Skuy, 2000). However, the present data go beyond these earlier studies by showing that the sex difference does not appear just on harder items or on those that come later in the test. If there is a difference between male and female performance on one type of item, but not on another, then the two types of items must be measuring or engaging at least partially discriminable processes or operations. The evidence from sex differences, therefore, confirms the conclusion based on factor analysis (Lynn, Allik, and Irwing), and experimental intervention (DeShon et al., 1995) that different items in Raven's Matrices must be assumed to be measuring several somewhat different processes: in addition to the distinction between perceptual and analytic items that has been widely recognised for some time, it is also important to draw a distinction between two types of analytic item.

We should not exaggerate the extent to which the Carpenter et al. (1990) rules are successful in their classification of Raven's items. As we noted in the introduction, in the Lynn et al. (2004) and Rushton and Skuy (2000) studies, although most of the items on which males outscored females were in Set E, and required the addition/subtraction or distribution of two rules, the last items in Sets C and D, which require no such rules, showed a similar pattern of male superiority. All items in Set C require the pairwise progression rule, and all those in Set D the distribution of three rule. In Lynn, Allik, and Irwing's confirmatory factor analysis, items C 12 and D 12 loaded strongly onto the same analytic factor as items E 7 to E 12, and presumably therefore engaged similar operations. The Carpenter et al. rules fail to illuminate the nature of this similarity. Equally, although the majority of items requiring addition/subtraction or distribution of two rules are classified by DeShon et al. (1995) as visuo-spatial, and all distribution of three items are classified as verbal–analytic, they classified some pairwise progression items as visuo-spatial, but we found no evidence of a sex difference on pairwise progression items.

The Carpenter et al. (1990) rules are, of course, merely descriptive. The finding that males obtain higher scores than females on items requiring addition/subtraction or distribution of two rules does not explain why they do better on these particular items. Since one of the few incontrovertible differences between males and females in IQ test performance is on tests of visuo-spatial ability or Gv, a possible explanation is that such items are somehow more "spatial" than others. This would agree with DeShon et al. (1995) classification—but not with that proposed by Lynn et al. (2004), whose visuo-spatial ability factor maps onto items requiring pairwise progression and distribution of three rules (and where they found no sex difference in performance), while their verbal–analytic reasoning factor maps onto addition/subtraction and distribution of two rules (where they did find that males outscored females). This perhaps highlights the danger of labelling factors simply by inspection of the tests that load onto them. Be this as it may, it has indeed been suggested that the occasionally observed male advantage on Raven's Matrices arises because of the spatial nature of the test (e.g., Colom & Garcia-Lopez, 2002; Colom, Escorial, & Rebollo, 2004; Flynn, 1998). There is certainly a reason to believe that Raven's is more likely than other tests of Gf or abstract reasoning to yield a male advantage. As we noted above, a small male advantage on RAPM or RPM is a common enough observation. On the other hand, Colom and Garcia-Lopez gave Spanish high-school students three different tests of Gf, RAPM, Cattell's CF test, and the inductive reasoning test from the PMA battery. Females obtained significantly higher scores than males on the PMA test; there was no sex difference on the Cattell test; but males significantly outscored females on RAPM. The authors argued that Raven's tests contained a spatial component and that "the male advantage on the Raven observed in the present investigation could come from its Gv ingredient" (p. 449).

If this is correct, and if the male advantage on RAPM is confined to items requiring addition/subtraction or distribution of two rules, it would seem to follow that it should be these items, and presumably these alone, that contain a Gv component. Although far from compelling, our results provide no support for this suggestion. It is true that males' Raven's scores were somewhat more highly correlated with their MRT scores than were females' (whose scores were somewhat more highly correlated with their DAT scores). This is certainly consistent with the idea that males were more likely than females to use spatial operations to solve Raven's and reinforces the general argument that different people may use different strategies to solve the same problem. But there was no suggestion that males' scores on addition/subtraction or distribution of two items were more highly correlated with their MRT scores than were their scores on other items (this was equally true for females). Our correlational evidence, therefore, fails to identify these items as peculiarly spatial. Set against this, the verbal overshadowing effect observed by DeShon et al. (1995) did suggest that such items were more spatial than those requiring the distribution of three rule (although it also suggested that some pairwise progression items were also spatial). We do not wish to place too much reliance on what is, in effect, a null result in our own data. It is certainly possible, for example, that a different spatial test might have given a different pattern of results: we chose to use the Vandenburg and Kuse (1978) MRT test because it reliably yields a substantial male advantage. But perhaps another spatial test, not requiring mental rotation, would have correlated more highly with these items than the pairwise progression or distribution of three items. All we can say is that the present data do not provide any evidence to show that addition/subtraction or distribution of two items have a stronger spatial component than other Raven's items. To that extent, it remains to be determined why they should be the only ones in RAPM to show a significant male advantage.

Acknowledgements

This research was supported by a grant from the U.K. Economic and Social Research Council to N.J. Mackintosh.

Appendix A

The 36 Raven's items used in the two specially constructed 18-item tests. Items E7, E8, E10, E11 and E12 are from Set E of RPM; all others are from RAPM. As explained in Table 1, we have provided our own classification for several RAPM items left unclassified by Carpenter et al. (1990).

Test 1		Test 2	
Raven's item	Carpenter et al. rule	Raven's item	Carpenter et al. rule
3	Pairwise	4	Pairwise
14	Pairwise	10	Pairwise
7	Add/Subtract	13	Distribution of 3
31	Distribution of 2	32	Distribution of 2
19	Add/Subtract	20	Add/Subtract
25	Pairwise	26	Pairwise
15	Add/Subtract	16	Add/Subtract
33	Add/Subtract	34	Distribution of 3
E7	Distribution of 2	E8	Distribution of 2
E10	Add/Subtract	E11	Add/Subtract
27	Distribution of 3	28	Distribution of 3
21	Distribution of 3	22	Distribution of 2
35	Distribution of 2	E12	Add/Subtract
17	Distribution of 3	18	Unclassified
29	Distribution of 3	30	Distribution of 2
23	Distribution of 2	24	Pairwise
8	Distribution of 3	9	Add/Subtract
6	Pairwise	5	Pairwise

References

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

Arthur, W., & Woehr, D. J. (1993). A confirmatory factor analytic study examining the dimensionality of the Raven's progressive matrices. *Educational and Psychological Measurement*, *50*, 471–478.

Burke, H. R. (1958). Raven's progressive matrices: A review and critical evaluation. *Journal of Genetic Psychology*, *93*, 199–228.

Carpenter, P. A., Just, M. A., & Schell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven's progressive matrices test. *Psychological Review*, *97*, 404–431.

Colom, R., Escorial, S., & Rebollo, I. (2004). Sex differences on the progressive matrices are influenced by sex differences on spatial ability. *Personality and Individual Differences*, *37*, 1289–1293.

Colom, R., & Garcia-Lopez, O. (2002). Sex differences in fluid intelligence among high school graduates. *Personality and Individual Differences*, *32*, 445–451.

Colom, R., & Lynn, R. (2004). Testing the developmental theory of sex differences in intelligence on 12–18 year olds. *Personality and Individual Differences*, *36*, 75–82.

Court, J. H. (1983). Sex differences in performance on Raven's progressive matrices: A review. *Alberta Journal of Educational Research*, *29*, 54–74.

DeSchon, 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.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, *43*, 95–103.

Flynn, J. R. (1998). Israeli military IQ tests: Gender differences small; IQ gains large. *Journal of Biosocial Science*, *30*, 541–553.

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

Hyde, G., & Trickey, G. (1995). *DAT for guidance. Manual*. London: The Psychological Corporation.

Jensen, A. R. (1980). *Bias in mental testing*. London: Methuen.

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

Lynn, R. (1998). Sex differences in intelligence: A rejoinder to Mackintosh. *Journal of Biosocial Science*, *30*, 529–532.

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

Lynn, R., Allik, J., Pullman, H., & Laidra, K. (2002). Sex differences on the progressive matrices among adolescents: Some data for Estonia. *Personality and Individual Differences*, *34*, 669–679.

Mackintosh, N. J., & Bennett, E. S. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence*, *31*, 519–531.

Mackintosh, N. J. (1998). Reply to Lynn. *Journal of Biosocial Science*, *30*, 533–539.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, *23*, 337–341.

Penrose, L. S., & Raven, J. C. (1936). A new series of perceptual tests: Preliminary communication. *British Journal of Medical Psychology*, *16*, 97–104.

Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural substrates of fluid reasoning: An MRI study of neocortical activation during performance of the Raven's progressive matrices test. *Cognitive Psychology*, *50*, 43–63.

Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. London: H.K. Lewis.

Rushton, J. P., & Skuy, M. (2000). Performance in Raven's Matrices by African and White university students in South Africa. *Intelligence*, *28*, 251–265.

Spearman, C. E. (1927). *The abilities of man*. London: Macmillan.

Spearman, C. E. (1946). Theory of the general factor. *British Journal of Psychology*, *36*, 117–131.

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.

Vandenburg, S. G., & Kuse, A. R. (1978). Mental rotations: Group tests of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*, 599–604.

Vernon, P. E., & Parry, J. B. (1949). *Personnel selection in the British Forces*. London: University of London Press.

Villardita, C. (1985). Raven's colored progressive matrices and intellectual impairment in patients with focal brain damage. *Cortex*, *21*, 627–634.