Educational and Psychological Measurement

http://epm.sagepub.com

Raven's Advanced Progressive Matrices: Norms for First-Year University Students and the Development of a Short Form

Douglas A. Bors and Tonya L. Stokes

Educational and Psychological Measurement 1998; 58; 382
DOI: 10.1177/0013164498058003002

The online version of this article can be found at:
http://epm.sagepub.com/cgi/content/abstract/58/3/382

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for *Educational and Psychological Measurement* can be found at:

Email Alerts: http://epm.sagepub.com/cgi/alerts

Subscriptions: http://epm.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations http://epm.sagepub.com/cgi/content/refs/58/3/382

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

RAVEN'S ADVANCED PROGRESSIVE MATRICES: NORMS FOR FIRST-YEAR UNIVERSITY STUDENTS AND THE DEVELOPMENT OF A SHORT FORM

DOUGLAS A. BORS AND TONYA L. STOKES
University of Toronto at Scarborough

Five hundred and six first-year university students completed Raven's Advanced Progressive Matrices. Scores on Set II ranged from 6 to 35 ($M = 22.17$, $SD = 5.60$). The first 12 items of Set II were found to add little to the discriminative power of the test. Exploratory and confirmatory factor analyses failed to confirm Dillon et al.'s two-factor solution and suggested that a single-factor best represented performance on Set II. A short-form of Set II, consisting of 12 items extracted from the original 36, was developed and found to possess acceptable psychometric properties. Although this short form differed considerably in content from the short form previously devised by Arthur and Day, the two short forms did not differ with respect to concurrent validity and predictive power.

Tests of inductive or analytic reasoning—what Cattell (1963) referred to as fluid intelligence—are said to estimate one's ability to solve problems without relying on an explicit base of knowledge derived from previous experience (Carpenter, Just, & Shell, 1990). Not only are such tests considered to be measures of specific forms of higher order cognitive abilities, but these tests are also considered to be among the best single indexes of general intelligence (Stough, Nettlebeck, & Cooper, 1993. For these reasons, tests of fluid intelligence have enjoyed wide use in both applied and research settings. Of such tests, Raven's Progressive Matrices tests (Raven, Raven & Court, 1991) have been among the most popular (Arthur & Woehr, 1993).

This article is specifically concerned with the Raven's Advanced Progressive Matrices (APM) (Raven, Court, & Raven, 1988), a version of the

Correspondence regarding this article should be addressed to Douglas A. Bors, Division of Life Sciences, University of Toronto at Scarborough, Scarborough, Ontario, Canada M1C 1A4; e-mail bors@scar.utoronto.ca.

Educational and Psychological Measurement, Vol. 58 No. 3, June 1998 382-398
© 1998 Sage Publications, Inc.

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009
from the SAGE Social Science Collections. All Rights Reserved.

matrixes intended for use with people of above average aptitude and designed to reliably differentiate among those in the top 25% of the population. During the past 15 years, the APM has probably been the instrument most widely used by researchers who are investigating the relations among various speed of information processing measures and intelligence (cf. Vernon, 1989). This particularly has been the case when undergraduate university students have been the population from which subjects have been drawn.

As with the other versions of Raven's Progressive Matrices, the APM has been found to yield reliable scores as a measure of general intelligence, and it correlated .74 with the full-scale Wechsler Adult Intelligence Scale (WAIS) and .75 with the Otis I.Q. (McLaurin, Jenkins, Farrar, & Rumore, 1973). The internal consistency of the APM has been found to be substantial, with split-half reliabilities ranging from .8 to .9 (Alderton & Larson, 1990; Arthur & Day, 1994; Paul, 1985). The test-retest stability has also been determined to be substantial (r = .83) (Bors & Forrin, 1995). Finally, together with the other above listed properties, the ease of administration to either individuals or groups has made the APM an ideal instrument for researchers.

Like the other Raven's matrixes tests, the APM is composed of a series of perceptual analytic reasoning problems, each in the form of a matrix. The problems involve both horizontal and vertical transformations: Figures may increase or decrease in size, and elements may be added or subtracted, flipped, rotated, or show other progressive changes in the pattern. In each case, the lower right corner of the matrix is missing and the subject's task is to determine which of eight possible alternatives fits into the missing space such that row and column rules are satisfied. The APM battery consists of two separate groups of problems. In both sets, the problems have been arranged such that each should be progressively more difficult than the preceding one. Set I consists of 12 problems that cover the full range of difficulty sampled in the Standard Progressive Matrices test (Raven et al., 1988). Standard timing for Set I is 5 min. This set is generally used as a practice test for those who will be completing Set II. Set II consists of 36 problems with a greater average difficulty than those in Set I. Set II can be administered in one of two ways: either with or without a time limit (40 min). Administering Set II without a time limit is said to specifically assess a person's capacity for clear thinking, whereas imposing a time limit is said to produce an assessment of intellectual efficiency (Raven et al., 1988).

Despite the growing number of studies using the APM to examine the relations between intelligence and performance on information processing tasks, many of which have drawn samples form North American university students, little has been published in the way of norms for this population. Unfortunately, in their study of 363 college students, Arthur and Woehr (1993) did not report any descriptive statistics. They did, however, report a small but statistically significant correlation between the APM and sex (–.09), with men slightly outperforming women. Paul's (1985) study of 300 Univer-

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

sity of California, Berkeley students (190 women, 110 men) was one exception in being explicitly normative, however. Tested under the untimed condition, the students' scores ranged from 7 to 36 with a mean of 27 ($SD = 5.14$). This was substantially higher than the mean of Raven's (1965) normative group ($M = 21.0$, $SD = 4.0$). In addition to an internal consistency (KR-20) of .83, Paul (1985) found a strong agreement between the rank order position of the problems as presented in the Set II booklet and the frequency with which the problems were solved ($r = .94$). A noted exception was Problem 13, which ranked as the 22nd most frequently solved problem. Corroborating the relation between the APM and measures of general intelligence, Paul reported that the APM correlated .69 with the Full Scale WAIS. Finally, Paul reported a statistically significant difference between the APM mean scores of men (28.40) and those of women (26.23); Hedges's *g* was .422.

In terms of its dimensionality, there have been conflicting findings and interpretations concerning the APM. Dillon, Pohlmann, and Hohman (1981), in a study of 237 secondary school students, concluded that the APM was dominated by two factors. One factor (reflected in Items 7, 9, 10, 11, 16, 21, 28, and 35) was interpreted to be an ability to solve problems whose solutions required adding or subtracting patterns; the other factor (reflected in Items 2, 3, 4, 5, 17, 26, 36) was interpreted to be an ability to solve problems whose solutions required detecting a progression in a pattern. Contrarily, from their study of 1,731 naval recruits, Alderton and Larson (1990) argued that a single-factor solution best described the APM. Additionally, they reported that their pattern of factor pattern coefficients was quite distinct from that of Dillon et al. Using 363 college students and confirmatory factor analytic techniques, Arthur and Woehr (1993) also failed to replicate Dillon et al.'s factor structure and concluded that a single-factor solution underlay response data.

Despite its popularity and its evident strengths, one disadvantage of the APM is its protracted administration time. As already mentioned, standard timing for Sets 1 and 2 is 5 min and 40 min, respectively. Additional time is also required before testing to provide instructions concerning the method of work. Thus, the total administration time may often last as long as 60 min. This makes it troublesome for researchers to administer any other psychometric tests or experimental tasks during the same testing session. A version of the test yielding reliable and valid scores that could be administered in 20 min or less would be of great utility to researchers.

Arthur and Day (1994) addressed the problem of lengthy administration time by creating an APM short form (Dillon Form) that they hoped would yield psychometric properties similar to that of the full-length APM. In their study, the original 36-item test was shortened to 12 items, requiring probably no more than 20 min to administer. To select the items for their short form, the original APM was divided into 12 equal sections of 3 items each, such that Items 1 through 3 comprised the first section, Items 4 through 6

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

comprised the second section, and so on. Of the three items in each section, the one with the highest item-total correlation was selected for use in the short form: Items 1, 4, 8, 11, 15, 18, 21, 23, 25, 30, 31, 35. Both the full-length APM and Arthur and Day's short form were subsequently administered to 246 university students. The short form's internal consistency (alpha = .65) was somewhat lower than that obtained from the full-length APM (alpha = .86); the correlation between the two forms was .66.

Although Arthur and Day (1994) have demonstrated that it is possible to develop a short form of the APM, it may still be possible to improve on the short form's psychometric properties. As seen in Paul's (1985, Table 1) and in Arthur and Day's (1994, Figure 1) reports, very few university students correctly answer fewer than 6 of the 36 items on the original APM so it is unlikely that Items 1 through 6 have much discriminatory power. Thus, the first two items on Arthur and Day's short form may essentially add noise to total scores. Additionally, possible redundancies in the form could have been identified by examining the inter-item correlations, but this apparently was not done. Finally, the issue of training and practice items to precede a short form has not been investigated. Do researchers need to administer all 12 practice items of Set II to subjects before administering a short form of Set II? Arthur and Day did not indicate the type of instruction or practice that preceded their administration of the short form.

The present research is intended to serve several purposes. To contribute further to the normative database for North American university students, we begin with a descriptive study of the performance of 506 students. Next, using both exploratory and confirmatory factor analyses, we examine the dimensionality of the APM and compare and contrast our findings with those of others. Using our findings, we then describe the construction of our new short form of the test and examine its performance. Given that the predictive utility of a test is limited by score reliability, and that the reliability of test scores is somewhat related to its length, the challenge is to produce a considerably shorter version of the APM without a substantial reduction in reliability. We then examine the effect that practice (Set I) may have on performance on the proposed new short form. Finally, we test the predictive utility of the new short-form APM by examining its relation to a speed of information-processing measure known to be correlated with the full-length APM.

Study 1

Subjects

The timed version of the full-length APM was administered to 506 students (326 women, 180 men) from the Introduction to Psychology course at the University of Toronto at Scarborough, most of whom subsequently

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

participated in various experiments involving information-processing tasks. Subjects ranged in age from 17 to 30 years, with a mean of 19.96 ($SD = 1.83$). Of the roughly 1,000 new students entering the University of Toronto at Scarborough each year, on average, 65% enroll in the introductory psychology course. Only a small proportion of these students (less than 10%) will complete a degree in psychology; most will major in disciplines in humanities, the social sciences, the biological sciences, or business administration. Approximately 25% will fail to complete a degree. Given that the course represents a large proportion of the students and broad samples from all disciplines, enrollment can be considered roughly representative of the first-year students.

Procedure

During the course of 2 years, small groups of subjects completed both Set I and Set II of the APM. Standard instructions were read aloud by the experimenter, and a standard timing of 5 minutes and 40 minutes were allotted for Set I and Set II, respectively. The scores for both sets were the total number of items completed correctly.

Results and Discussion

Scores on Set I ranged from 2 to 12 with a mean of 9.40 ($SD = 1.76$). Given that Items 1 and 2 of Set I are used for instructional purposes, actual scores ranged from 0 to 10 ($M = 7.40$). Table 1 provides the by-item frequencies with which each alternative was chosen. As can be seen from the table, some of the distracters were rarely if ever chosen; this was particularly true for the less difficult items (Items 3 to 7). As can also be seen from Table 1, although the items tended to increase in difficulty, several items appear to be out of order. This is true even for the more difficult items. With respect to the easier items (Items 3, 4, 5, 6, and 7), order appears to be irrelevant for this student population: Errors are few and appear to be fairly random. For the more difficult items, a more appropriate ordering, however, might be 8, 10, 9, 12, 11. As would be expected for a test designed to sample all levels of ability, the distribution of scores on Set I was negatively skewed, indicating that Set I was relatively easy for these university students. Less than 1% of those tested made more than seven errors, whereas 72% had scores of nine or greater. Finally, the performance of the men ($M = 9.54$, $SD = 1.72$) was slightly superior to that of the women ($M = 9.33$, $SD = 1.78$) but not significantly so, $F(1, 505) = 1.72$, $MSE = 1546.55$, $p > .05$. Hedges's g standardized difference was .120.

The scores on Set II for the 506 students ranged from 6 to 35 with a mean of 22.17 ($SD = 5.60$). This performance is somewhat higher than that of

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

Table 1
Advanced Progressive Matrices (APM) (Set I): Response Frequencies by Item

Item	Response Number									
Number	1	2	3	4	5	6	7	8	NR	Passed
1	0	0	0	0	0	0	0	506[a]	0	100
2	0	0	0	506[a]	0	0	0	0	0	100
3	4	3	6	3	482[a]	0	4	3	0	95
4	494[a]	4	0	1	5	2	0	0	0	98
5	4	484[a]	2	0	3	4	6	3	0	96
6	10	21	3	4	452[a]	0	3	13	0	89
7	9	0	0	6	5	479[a]	2	1	4	95
8	10	23	384[a]	2	29	13	7	14	24	76
9	9	6	31	22	7	20	304[a]	53	54	60
10	10	4	10	11	0	21	10	327[a]	113	65
11	58	10	3	19	13	30	154[a]	13	206	30
12	13	3	6	4	10	172[a]	6	21	271	34

Note. NR = no response.
a. correct response.

Raven's (1962) normative group but considerably lower than Paul's (1985) University of California, Berkeley sample. Unfortunately, because Paul did not provide details of his sampling procedures, we do not know if his 300 students were drawn from the general population of first-year students, as was the our sample. Furthermore, where Paul administered the test untimed, we administered it with the 40-minute time limit, and this, in part, may be responsible for the difference between the means. Like Paul, however, we found the difference between the mean total scores of men (M = 23.00) and women (M= 21.68), F(1, 504) = 7.11, MSE = 30.93, $p < .05$, to be statistically significant. Hedges's g was .559. Also like Paul, we cannot rule out sampling error as an explanation. Typically, in the introductory psychology class at the University of Toronto at Scarborough, women volunteer to participate in psychological research at a greater rate than do men. In the present case, although men comprised 45% of the introductory course over the period in question, only 36% of our sample were men. In comparison to their women cohorts, it is possible that fewer men from the lower tail of the distribution of APM scores volunteer.

As was the case with Set I, many of the distractors were rarely if ever chosen. This was particularly the case with the first 14 items. And again, although the items as ordered tended to increase in difficulty, several items appear to be out of order. Again, this is also true for the more difficult items. With respect to the easier items (Items 1 to 9), order again appears to be irrelevant for this population in that errors are few and essentially random.

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

Table 2
Item-Total Correlations

Item Number	Item-Total Correlation	Item Number	Item-Total Correlation	Item Number	Item-Total Correlation
1	.29	13	.31	25	.33
2	.22	14	.31	26	.32
3	.38	15	.38	27	.36
4	.33	16	.41	28	.37
5	.25	17	.27	29	.34
6	.22	18	.39	30	.39
7	.22	19	.37	31	.39
8	.31	20	.28	32	.26
9	.32	21	.49	33	.25
10	.44	22	.41	34	.37
11	.34	23	.32	35	.27
12	.41	24	.36	36	.06

This suggests that these items provide little in the way of discriminative power for this population, the one possible exception being Item 4, which might be better placed in the 10th position. The inconsequence of these easier items is further suggested by the fact that only 2.5% of the students answered fewer than 10 of the 36 items correctly. This suggests that an abbreviated version of the test is possible.

With respect to the more difficult items (Items 10 to 36), several items appear to be clearly out of order. A more appropriate ordering for the more difficult items might be 11, 4, 10, 12, 14, 15, 16 17, 19, 13, 18, 21, 20, 23, 22, 25, 26, 24, 27, 30, 28, 31, 29, 32, 33, 34, 35, 36. Table 2 provides the item-total score correlations, with the item in question removed from the total score. For the population from which we sampled, with the exceptions of Items 3 and 34, most of the discriminatory power appears to reside in middle-range items. In particular, save for the single exception of Item 3, the first 12 items contribute the least to the test's discriminative power. When the test was sequentially divided into three equal parts (Items 1 to 12, Items 13 to 24, and Items 25 to 36) and each section was scored separately, the resulting correlations with total score were .75, .88, and .78 for the first, second, and third sections, respectively.

The internal consistency of the 36 items, based on Cronbach's alpha, was .84. This finding is consistent with that of Arthur and Day (1994), who also reported an alpha of .84 for their sample of 202 university students. Finally, the correlation between scores on Set I and scores on Set II was .53, indicating that performance on the 12 practice items moderately predicts performance on the 36 test items.

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

Set II Factor Structure

For a set of dichotomously scored items, particularly a set with substantially different levels of difficulty as is the case with the APM, the results of factor analytic techniques using a matrix of phi coefficients are often spurious (Gorsuch, 1983). One solution to this problem, and the one employed here, is to use a matrix of tetrachoric inter-item correlations (Arthur & Woehr, 1993). A principal components analysis of the tetrachoric correlation matrix computed for the 36 items of Set II produced 12 factors with eigenvalues greater than 1. The first three factors had eigenvalues of 12.04, 3.28, and 2.19, respectively, and accounted for 48.63% of the variance. All other eigenvalues were less than 2.00. As would be expected, all variables were positively correlated with the first factor.

Dillon et al. (1981) reported a two-factor solution, rotated using an orthosim orthogonal solution. The orthosim rotation produces solutions that are similar to those produced by a standard varimax rotation (Bentler, 1977). The factor pattern coefficients for both our data and those reported by Dillon et al. are presented in Table 3.

As can be seen from the Table 3, the factor pattern coefficients derived from the two data sets are quite different. Save for the six or seven most difficult items, the items from our data set correlated more consistently with one factor than do the items from Dillon et al.'s (1981) data set. Specifically, concerning the 15 items Dillon et al. identified with their two-factor solution, all items in our analysis with the exceptions of Item 35 and Item 36 correlated most strongly on a single factor. Thus, little support for Dillon et al.'s two-factor solution is provided by the exploratory factor analysis of our data.

To further test Dillon et al.'s (1981) two-factor solution, the 15 items they identified, and only those items, were used as input for confirmatory factor analyses using EQS (Bentler, 1995). It is possible that when only these 15 items are considered, they would segregate differently than they did when all 36 items were considered in the exploratory analysis. Furthermore, any new pattern of segregation might be more consistent with that reported by Dillon et al. Three models were evaluated: a single factor solution, a solution with two independent factors, and a solution with two correlated factors. As identified by Dillon et al., in both of our two-factor models, Items 2, 3, 4, 5, 17, 26, and 36 were associated with one factor and Items 7, 9,10,11,16, 21, 28, and 35 were associated with the other factor. A summary of the results is reported in Table 4.

As with any χ^2 test of fit, the greater the χ^2 value, the poorer the fit between the model and the data, and statistical significance indicates that a model fails to provide an adequate statistical fit. As can be seen from Table 4, none of the three models tested provided an adequate fit for the inter-item correlations. The comparative fit indexes suggest that both the one factor solution and the two-factor (correlated) solution approached adequacy. Both the single-factor

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

Table 3
Factor Pattern/Structure Coefficients for Dillon, Pohlmann, and Hohman's (1981) Data and for the Present Data

	Dillon et al.'s Data					Present Data					
	Factor			Factor			Factor			Factor	
Item Number	I	II	Item Number	I	II	Item Number	I	II	Item Number	I	II
1	.14	.52	19	.21	.15	1	.65	.29	19	.51	.17
2	.26	.82	20	.23	.56	2	.60	.22	20	.42	.05
3	.07	.74	21	.60	.26	3	.81	.17	21	.75	.10
4	.17	.79	22	.59	.24	4	.51	.36	22	.53	.24
5	.30	.63	23	.56	.12	5	.50	.10	23	.36	.31
6	.49	.41	24	.50	.31	6	.65	–.10	24	.54	.18
7	.77	.16	25	.43	.40	7	.53	.11	25	.38	.19
8	.45	.49	26	–.04	.59	8	.60	.12	26	.39	.21
9	.85	.18	27	.27	.32	9	.58	.29	27	.52	.24
10	.67	.38	28	.67	.16	10	.73	.25	28	.40	.26
11	.60	.44	29	.38	.47	11	.65	.20	29	.36	.44
12	.47	.41	30	.20	.41	12	.70	.18	30	.34	.59
13	.55	–.04	31	.52	.52	13	.56	–.08	31	.34	.56
14	.30	.43	32	.42	.43	14	.63	–.04	32	.18	.55
15	.48	.30	33	.28	.35	15	.64	.06	33	.13	.67
16	.61	.34	34	.42	.40	16	.51	.35	34	.27	.83
17	.21	.63	35	.74	.12	17	.32	.29	35	.23	.62
18	.48	.28	36	.17	.79	18	.56	.17	36	–.25	.74

model, χ^2 (1) = 106.37, $p < .05$, and the two-factor (correlated) model, χ^2 (2) = 109.69, $p < .05$, were better fits than the two-factor (independent) model. The correlation between the two factors in the two-factor (correlated model) was .86 ($p < .05$), however, denoting very little difference in the two factors and suggesting the presence of a single higher order factor. Furthermore, the difference between the single-factor model and the two-factor (correlated) model was nonsignificant, χ^2 (1) = 3.32. With respect to a first-year university population, in light of the results of both the exploratory and confirmatory factor analyses, we conclude that a single-factor solution best represents Set II of the APM.

Development of a Short Form

Because all APM items share the same format and because we have accepted, at least tentatively, the idea that a single factor underlies performance on the APM, there was no need to consider sampling items from various subsets when we constructed our short-form APM; discriminability was the only issue of concern to us. Thus, we began by rank ordering the items by their item-total correlations found in Table 4. Next, we examined the inter-item correlations to remove any redundancies. An item may have a

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

Table 4
Confirmatory Factor Analysis: Goodness of Fit for Three Models

Model	χ^2	*df*	CFI	BBNFI
One factor	133.04*	89	.901	.883
Two factors (independent)	239.41*	88	.667	.603
Two factors (correlated)	129.72*	87	.906	.887

Note. χ^2 = chi-square goodness-of-fit value; *df* = degrees of freedom; CFI = Comparative Fit Index (Bentler, 1988); BBNFI = Bentler-Bonett Nonnormed Fit Index (Bentler, 1988).
*$p < .05$.

relatively high item-total correlation but also be substantially correlated with another item with a similarly high item-total correlation, thus adding little to the predictive power of the test. As a result, Items 3, 10, 12, 15, 16, 18, 21, 22, 28, 30, 31, and 34 were selected for our new short-form APM.

The difference between the set of items that we have selected and the set chosen by Arthur and Day (1994) is substantial. Only five items coincide in the two short forms. The primary reason for the differences appears to be related to Arthur and Day's (1981) strategy of sequentially dividing the test into 12 sections of three items each and then selecting one item from each section. This meant, by necessity, that Arthur and Day (1994) would select three items from the first quarter of the test (Items 1, 4, and 8). Using no such restriction, we selected only one item (Item 3) from this subset.

Short-Form Performance

Scores for our new short-form APM based on the 12 selected items ranged from 0 to 12 (M = 7.01, SD = 2.56). When Arthur and Day's (1994) short form was applied to our data, scores ranged from 1 to 12 (M = 7.49, SD = 2.30). The difference between the two means was statistically significant ($t_{505} = 7.81, p < .001$), Hedges's $g = .200$, indicating that our form was more difficult than that of Arthur and Day's. This undoubtedly related to Arthur and Day's item selection strategy. In comparison to Arthur and Day's form, the somewhat greater variance, reduced skewness, and flatter distribution found for our form is principally due to the larger percentage of participants obtaining scores of 3 or less on our short form. With Arthur and Day's form, 5% had scores of 3 or less, whereas with our form, 10% had scores of 3 or less.

The internal consistency of our form, based on Cronbach's alpha, was .73. This is consistent with that reported by Arthur and Day (1994) for their short-form APM (alpha = .72). Although our alpha is somewhat lower than the .84 we found for full-length APM scores, the internal consistency remains high, given the two thirds reduction in the number of items.

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

The correlation between our new short form and the full-length APM was .92 ($p < .001$). This is marginally stronger than the correlation (.90) between the short-form and the full-length APMs reported by Arthur and Day (1994). Additionally, when we applied Arthur and Day's short form to our data, the resulting correlation with the full-length APM was .89 ($p < .001$), again only marginally weaker than the correlation between our short form and the full-length APM. Although not statistically significant, the slight difference between the correlations is likely attributable to the somewhat greater variance produced by our version.

Study 2

The first purpose of this second study was to determine the effect of the 2 instructional items and 10 practice items (Set I) of performance on our new short-form APM. Given that the primary motive for developing a short-form APM was to reduce the time required to administer the battery, reducing the number of necessary practice items would be of additional benefit. Should there be no difference between subjects who received all 12 items in Set I and those who received only the two instructional items (Item 1 and Item 2) then it may more safely be assumed that the short-form (with only the two instructional items from Set I) will produce total score distributions with properties similar to those found in Study 1, thereby allowing further time to be saved. The second purpose of this study was to consider the test-retest reliability of scores on the new short-form APM. A crucial characteristic of any measure of individual differences with predictive power is stability over time. The final purpose was to examine how the correlation between our short-form APM and another group-administered intelligence test compares with the correlation between the full-length APM and the same test.

Participants

Participants in Group 1 were 53 volunteer second-year students at the University of Toronto at Scarborough who ranged in age from 18 to 25 years ($M = 21.79$, $SD = 1.77$). Thirty-six of the participants were women, and 17 were men. Participants in Group 2 were 41 volunteer first-year and second-year students from the same university who ranged in age from 18 to 24 years ($M = 21.84$, $SD = 1.57$). Twenty-six of the participants in Group 2 were women, and 15 were men.

Procedure

Group 1 participants were randomly assigned to one of two conditions. All participants in Group 1 completed the new short-form APM. Half of them

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

(n = 26) were administered all 12 practice items from the original Set I (the 2 instructional items and the 10 practice items) prior to the new short form of Set II. The other half of the participants in Group 1 (n = 27) were only administered the two instructional items (Items 1 and 2) from Set I prior to the short form for Set II. On completion of the short-form APM on the first occasion, participants were administered the Abstraction subtest of the Shipley Institute of Living Scale (Shipley) (Zachary, 1991), with standard instructions and timing (10 minutes). Participants in Group 2 were administered the full-length APM and the Abstraction subtest of the Shipley, again using standard instructions and timing. After a period of 2 weeks, 39 of the 53 participants in Group 1 were retested with our short-form APM in the same manner as they had been on the first occasion.

Results and Discussion

The overall mean on the new short-form APM was 7.39. Although the mean for those Group 1 participants who received all 12 items of Set I (M = 7.63, SD = 2.47) was somewhat greater than that for those participants who received only the first two items of Set I (M = 7.15, SD = 2.34), the difference was not statistically significant, $F(1, 51) < 1$. Hedges's g was .199. This indicates that the effect of Items 3 through 12 on Set I was minimal and that only the first two instructional items need be administered prior to the new short form of Set II.

For those 38 participants who were retested 2 weeks later, the test-retest reliability of the new short-form APM was .82. The test-retest reliabilities for those who received all 12 items of Set I (n = 20) and for those who received only the first two instructional items (n = 18) were .81 and .84, respectively. These findings can be considered more than adequate when compared to the test-retest correlations found for the full-length APM (.83) reported by Bors and Forrin (1995). Furthermore, the similar consistency for the two subgroups again suggests that only the first two instructional items from Set I need be administered prior to the new short form of Set II.

Scores on the Abstraction subtest of the Shipley for Group 1 ranged from 12 to 20 (M = 17.06, SD = 1.95). Scores for Group 2 ranged from 10 to 20 (M = 16.67, SD = 2.31). Scores on the full length APM for Group 2 ranged from 12 to 30 (M = 23.76, SD = 5.28). The correlation between the full-length APM scores for Group 2 and their scores on the Abstraction subtest of the Shipley was .73. The correlation between the short-form APM scores for all subjects in Group 1 and their scores on the Abstraction subtest of the Shipley was .61. The difference between these correlations was nonsignificant, Z = 1.10. The short-form APM Abstraction subtest correlations for those who received all 12 items from Set I and those who received only the first two items were .60 and .74, respectively. It would appear that the concurrent

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

validity of the APM scores is not substantially reduced when our short-form APM is administered after only the two instructional items from Set I.

Study 3

The purpose of Study 3 was to compare the strength of the correlation between scores on our short-form APM and performance on a simple information-processing task with the strength of the correlation between scores on the full-length APM and the same information-processing task. To do so, we used a task—inspection time (IT)—whose relation to scores on the full-length APM is well established. IT is a speed of information-processing paradigm initially developed by Vickers, Nettlebeck, and Willson (1972) to estimate visual encoding time. Authors of comprehensive reviews of the literature have concluded that there is a stable moderate correlation (approximately –.50) between IT and scores on intelligence tests (IQ), particularly those instruments, like the APM, that are said to measure performance IQ (Kranzler & Jensen, 1989; Nettlebeck, 1987). Because we have reduced the APM to one third of its length, we must assume that some discriminative and predictive power will be lost. The question is how much? The findings of Study 1 and Study 2 suggest that the losses in predictive power that accompany use of our short form may be small.

Participants

Forty-five volunteers (19 men and 26 women), with normal or corrected-to-normal vision, were recruited from an introductory psychology class at the University of Toronto at Scarborough. Participants ranged in age from 18 to 27 years ($M = 22.2$, $SD = 2.01$).

IT

The IT stimulus, taken from Vickers et al. (1972), consisted of two vertical lines (2.8 and 3.8 cm in length), 0.8 cm apart, and connected at the top by a 1.8 cm horizontal line. On half of the trials, the longer of the two vertical lines appeared on the left side of the stimulus display, and on the other half of the trials, the longer line appeared on the right side. The backward mask, taken from Nettlebeck and Rabbitt (1992), had the same general appearance as the stimulus display, except that the two vertical lines were both 4.6 cm long and thickened in the center in a manner resembling lightning bolts.

Psychometric Test

Both complete sets of the APM were administered to the participants using standard timing, 5 min and 40 min, respectively. A full-length score, a score

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

for our short-form, and a score for Arthur and Day's (1994) short-form were derived from the 36 items for Set II.

Procedure

After the participants individually completed the APM, the IT task was explained and administered. The method of constant stimuli was used so that IT (using a 95% accuracy criteria) could be interpolated from the commutative normal ogive. Each participant was seated in a dimly lighted room and positioned approximately 30 cm from a computer monitor. Each trial began with a warning sound and the display of a focal point positioned on the screen such that it would be midway between the two vertical lines and 18 cm below the horizontal line of the stimulus display. Following a 1 s foreperiod, the focal point was removed and the stimulus displayed for 1 of 10 exposure durations (14 to 266 ms with 28 ms intervals). At the conclusion of the duration, the backward mask replaced the stimulus display. Testing consisted of 300 trials, 30 trials at each duration. The participant's task was to determine on which side of the stimulus display the longer of the vertical lines appeared. A short training session of 10 trials, one at each exposure duration, preceded the 300 test trials.

Results and Discussion

Total accuracy on the IT task ranged from 140 to 282 ($M = 259.69$, $SD = 21.58$) out of 300. Derived IT scores corresponding to a 95% accuracy criterion ranged from 54.09 to 224 ms ($M = 110.62$, $SD = 45.43$). These results are within the range typically reported by other researchers (Nettlebeck, 1987). The scores on the full-length APM ranged from 10 to 33 ($M = 24.20$, $SD = 5.68$). When only those 12 items used in our short form were analyzed, scores ranged from 2 to 11 ($M = 7.88$, $SD = 2.45$). The correlations between total accuracy, IT, full-length APM and our short-form APM are found in Table 5. Again, consistent with previous findings, IT and full-length APM scores were moderately correlated. As expected, full-length and short-form APM scores were substantially correlated. Both full-length and our short-form APM scores were moderately correlated with IT. Although, as expected, the correlation between our short-form APM and IT was weaker than that between the full-length APM and IT, it remained moderately strong and statistically significant, even with the relatively small number of participants in the study.

The scores derived from the 12 items used in Arthur and Day's (1994) short form ranged from 3 to 12 ($M = 8.40$, $SD = 2.34$). As expected from the results of Study 1, scores on their short form were slightly higher than those on our form. Again, as expected, scores on Arthur and Day's short form were

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

Table 5
Correlations Between Inspection Time Measures and Advanced Progressive Matrices (APM) Scores

APM		Our Short Form	Arthur and Day's Short Form	IT
Our short form	.88	1.00		
Arthur and Day's short form	.87	.88	1.00	
IT	–.48	–.42	–.42	1.00

Note. APM is the full-length APM. Our short form is our proposed short-form APM. Arthur and Day's short form is Arthur and Day's (1994) proposed short-form APM. IT is Vickers, Nettlebeck, and Willson's (1972) inspection time task.

strongly correlated with both the full-length scores and those on our short form. As seen in Table 5, the correlation between Arthur and Day's short form and IT was identical to that between our short form and IT. Considering the issue of Type II errors, it appears that researchers would not need to substantially increase the number of subjects they tested with either short-form to equal the power of a study employing the full-length APM.

General Discussion

In Study 1, 506 first-year university students were administered both Set I and Set II of Raven's APM with standard instructions and timing. With respect to Set II, the first 9 items presented little challenge, with accuracy rates ranging from 86% to 95% and with no particular order of difficulty emerging. It was the middle 12 items (Items 13 to 24) that were found to have the greatest discriminatory power. The first third of Set II, particularly the first 9 items, may do little more than identify the lower end of the distribution of university students. Only 18 of the 506 students tested had total scores below 12. Both exploratory and confirmatory factor analyses suggested that a single-factor solution was appropriate and specifically failed to support Dillon et al.'s (1981) two-factor solution.

In an attempt to reduce administration time and to perhaps improve on a short-form APM previously developed by Arthur and Day (1994), using the above analyses, a short form of Set II was developed consisting of Items 3, 10, 12, 15, 16, 18, 21, 22, 28, 30, 31, and 34. Both the form's internal consistency and the form's correlation with full-length scores were found to be satisfactory. Although our short form was significantly more difficult than the of Arthur and Day's short form, the correlations with full-length scores did not significantly differ.

Study 2 examined the effect of Set I on performance on our short form for Set II, the correlation between our short form and another group intelligence test (the Abstraction subtest of the Shipley), and the test-retest reliability of

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

our short form. It was found that, beyond the first two instructional items, completion of Set I had no statistically significant effect on performance on our short form. This permits a further reduction in administration time. Furthermore, using our short form did not appear to result in any substantial loss of concurrent validity. Whereas the correlation between full-length APM scores and the Abstraction subtest of the Shipley was .73, the correlation between scores on our short form and the Shipley subtest was .61, a statistically nonsignificant difference. The test-retest reliability of our short form was found to be more than acceptable ($r = .82$, $n = 38$).

In Study 3, scores on the full-length APM, our short form, and Arthur and Day's (1994) short form were correlated with performance on Vickers et al.'s (1972) IT task, a frequently used speed of information-processing task known to be moderately correlated with full-length APM scores. Scores on the full-length APM correlated –.48 with IT. Scores from our short form and Arthur and Day's short form both correlated –.42 with IT. Both short forms evidently perform adequately as substitutes for the full-length version of the APM (Set II). The modest psychometric advantages our short form may have over Arthur and Day's short form did not yield any difference in predictive power.

In conclusion, for most research purposes, when sampling from a first-year university population, using only the first two items of Set I for instruction together with either our or Arthur and Day's (1994) short form of Set II will allow for substantial time savings without seriously compromising the reliability or the validity of APM scores.

References

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

Arthur, W., & Day, D. (1994). Development of a short form for the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement*, *54*, 395-403.

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. (1977). Factor simplicity index and transformation. *Psychometrika*, *42*, 277-295.

Bentler, P. M. (1995). *EQS: Structural equation program manual*. Encino, CA: Multivariate Software Inc.

Bors, D. A., & Forrin, B. (1995). Age, speed of information processing, recall, and fluid intelligence. *Intelligence*, *20*, 229-248.

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*, *98*, 404-431.

Cattell, J. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1-22.

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

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Kranzler, J. H., & Jensen, A. R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence, 13*, 329-347.

McLaurin, W., Jenkins, J., Farrar, W., & Rumore, M. (1973). Correlations of IQ on verbal and non-verbal test of intelligence. *Psychological Reports, 33*, 821-822.

Nettlebeck, T. (1987). Inspection time and intelligence. In P. A. Veron (Ed.), *Speed of information-processing and intelligence* (pp. 295-346). Norwood, NJ: Ablex.

Paul, S. M. (1985). The Advanced Raven's Progressive Matrices: Normative data for an American university population and an examination of the relationship with Spearman's *g*. *Journal of Experimental Education, 54*, 95-100.

Raven, J. C., Court, J. H., & Raven, J. (1988). *Manual for Raven's Progressive Matrices and Vocabulary Scales* (Section 4). London: H. K. Lewis.

Raven, J. C., Raven, J., & Court, J. H. (1991). *Manual for Raven's Progressive Matrices and Vocabulary Scales* (Section 1). Oxford, UK: Oxford Psychologists Press.

Vernon, P. A. (1989). *Speed of information-processing and intelligence*. Norwood, NJ: Ablex.

Vickers, D., Nettlebeck, T., & Willson, R. J. (1972). Perceptual indices of performance: The measurement of "inspection time" and "noise" in the visual system. *Perception, 1*, 263-295.

Zachary, R. A. (1991). *Shipley Institute of Living Scale: Revised manual.* Los Angeles: Western Psychological Services.

Downloaded from http://epm.sagepub.com at LUND UNIVERSITY LIBRARIES on December 25, 2009