

# Educational and Psychological Measurement

<http://epm.sagepub.com/>

---

## The 20-Minute Version as a Predictor of the Raven Advanced Progressive Matrices Test

Ronald Hamel and Verena D. Schmittmann  
*Educational and Psychological Measurement* 2006 66: 1039  
DOI: 10.1177/0013164406288169

The online version of this article can be found at:  
<http://epm.sagepub.com/content/66/6/1039>

---

Published by:



<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/content/66/6/1039.refs.html>

---

# The 20-Minute Version as a Predictor of the Raven Advanced Progressive Matrices Test

Ronald Hamel  
Verena D. Schmittmann  
*University of Amsterdam*

The Raven Advanced Progressive Matrices Test (APM) is a well-known measure of higher order general mental ability. The time to administer the test, 40 to 60 minutes, is sometimes regarded as a drawback. To meet efficiency needs, the APM can be administered as a 30- or 40-minute timed test, or one of two developed short versions could be used. In this study, the 20-minute timed version of the APM is compared to the untimed APM as a measure of intellectual ability in 1st-year psychology students. This 20-minute timed version proves to be an adequate predictor of the untimed APM score.

**Keywords:** *intelligence measures; Raven Advanced Progressive Matrices; test administration; group testing; 20-minute timed version*

The Raven Progressive Matrices Test (RPM) and the Raven Advanced Progressive Matrices Test (APM; Raven, Raven, & Court, 1993) are widely used to measure problem-solving ability or *eductive ability* (Raven et al., 1993), fluid intelligence (Cattell, 1963), and analytic intelligence (Carpenter, Just, & Shell, 1990; cf. g; Spearman, 1927). As Carpenter et al. (1990) showed, the RPM measures the common ability to “decompose problems into manageable segments and iterate through them, the differential ability to manage the hierarchy of goals and subgoals generated by this problem decomposition, and the differential ability to form higher level abstractions” (p. 429).

The RPM and APM are used in daily practice as well as in research settings. The time needed to administer the tests is often regarded as a drawback: 30 or 40 minutes in the timed version for the APM and even longer in the untimed version, plus 20 minutes

---

**Authors' Note:** We thank Jan Elshout for his valuable advice at the beginning of the study; Emoke Jakab for her assistance; and Bob Bermond, Conor Dolan, Martin Elton, Jeroen Raaijmakers, and Jelte Wicherts for their comments on earlier versions of the article. Correspondence concerning this article should be addressed to Ronald Hamel, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, the Netherlands; e-mail: [r.hamel@uva.nl](mailto:r.hamel@uva.nl).

for instruction and practice. Raven, Raven, and Court (1998) reported norms for the APM Set II with time limits of 30 and 40 minutes. In an attempt to reduce the time needed to obtain valid and reliable scores on the APM, Arthur and Day (1994) and Bors and Stokes (1998) developed short versions of the APM. Both short versions consist of 12 items selected from the 36 items in Set II. Arthur and Day (1994) selected 12 items by dividing the APM in 12 sections of 3 items and choosing from each section the item with the highest item-total correlation. Bors and Stokes (1998) selected a set of 12 items by rank-ordering the items by their item-test correlations, with the item in question removed from the total score, and by removing from that list 24 items on the basis of interitem correlations to remove redundancies. The overlap of both short versions consists of 5 items. Arthur, Tubre, Paul, and Sanchez-Ku (1999) reported norms for the short version proposed by Arthur and Day. However, 12 items selected from the 36-item-long APM might represent a task that differs from the original APM. As a consequence, the validity of the short version as a measure of problem-solving ability or eductive ability might be affected.

In the APM, the level of difficulty of the items increases gradually. As a consequence of the selection of 12 items out of 36, the increase in difficulty of the 12 items remains the same as the increase in difficulty of the 36 items of the whole APM, but the steps between items are greater (the increase is steeper). The validity of the APM as a power test bears quite heavily on learning from experience during the test (Raven et al., 1993); therefore, these short versions might differ from the APM in a qualitative way that may not be intended. There remain fewer instances to learn from experience or practice (12 instead of 36), while the differences in difficulty between these instances are greater.

The APM could also be administered with a time limit, as a speed test. In this case it assesses *intellectual efficiency* (Raven et al., 1993), while practice and experience with previous items remain to play a role as in the untimed APM. Whereas the original, untimed APM is considered a unidimensional test (Raven et al., 1998) a timed version of the APM might additionally involve a speed factor as well. Although there exist norms for timed APM versions of 30 minutes and 40 minutes (Raven et al., 1998), the question remains if timed APM scores might be biased by a confounding speed factor. The characteristics of such a bias have not yet been investigated.

Another way to arrive at a short version might be to administer a timed version and an untimed version of the APM and to investigate how well scores on the timed version and of subsequent parts of the APM corresponding with increasing time intervals predict scores on the untimed APM. Our study investigates how well scores on the APM after 20 minutes, after 30 minutes, and after 40 minutes, respectively, predict untimed completion of the test and how well scores on a 20-minute timed version predict untimed completion of the test.

There is a difference between the short versions of Arthur and Day (1994) and of Bors and Stokes (1998), on one hand, and our approach, on the other. The task of someone doing the short versions of Arthur and Day and of Bors and Stokes is different from the first 20 minutes of the whole APM, because their items are samples from the

APM. The task of our participants is identical to the first 20 minutes of the whole APM, because it consists of all items of the APM.

The purpose of this study is to evaluate the prediction of APM scores on the basis of scores on a 20-minute version of the APM by comparing the participants' score after 20 minutes, 30 minutes, 40 minutes, and longer (as long as needed to complete the test if longer than 40 minutes).

## Method

The participants were drawn from a group of 542 first-year undergraduate psychology students in the Department of Psychology at the University of Amsterdam, who by participating fulfilled a course requirement. Of these students, 492 (91%) reported Dutch as their first language. The parents of the 542 students were either born in the Netherlands (82%), in another European country (7%), in Surinam and the Netherlands Antilles (4%), or elsewhere (7%).

First, the APM was administered to 51 students (38 women, 13 men). These participants' age ranged from 17 to 28 years, with a mean of 19.57 years ( $SD = 1.98$ ).

Two months later, the APM was administered to the 51 students again, but this time as a speed test; the allotted time was 20 minutes. At the same occasion, this version of the APM was administered to 397 other students (273 women, 124 men). These participants' age ranged from 17 to 30 years, with a mean of 20.37 years ( $SD = 2.57$ ). In the course of this study, these students completed the test once.

All 51 plus 397 participants were also administered six other ability tests (Elshout, 1976), based on different factors in Guilford's (1967) Structure of Intellect model. Each year, for more than three decades, all 1st-year psychology students at the University of Amsterdam take these tests as a course requirement. Taken together they are considered as an intelligence test. *Conclusions* consists of linear syllogisms, items like "A > C, C > B, what is the relation between A and B?" It measures Cognition of Semantic Systems (CMS) and Evaluation of Semantic Implications (EMI). *Number Series* measures the ability to recognize the "system" in a series of numbers or symbols: Cognition of Symbolic Systems (CSS). *Arithmetic Speed* measures the ability to apply simple symbolic rules, addition, subtraction, multiplication, and division: Convergent Production of Symbolic Implications (NSI). *Verbal Analogies* consists of items like "foe : hatred = friend : . . . ?" It measures Cognition of Semantic Relations (CMR). *Vocabulary* is a verbal ability test and measures knowledge of the meaning of words: Cognition of Semantic Units (CMU). *Embedded Figures* measures the ability to single out one figure in a complex line pattern: Convergent Production of Figural Transformations (NFT). The sum of the standardized scores on these six tests is used as an overall measure of intelligence. See Table 1 for details regarding the scoring of the six tests and their sum.

**Table 1**  
**Scoring of the Six Tests**

	Number of Items	Possible Range	Time Limit in Minutes	Guttman's Split Half
Conclusions	40	0 through 40	8	.93
Number Series	30	0 through 30	12	.83
Arithmetic Speed	90	0 through 90	4	.97
Verbal Analogies	40	0 through 40	5 + 5	.71
Vocabulary	40	0 through 40	10	.68
Embedded Figures	32	0 through 32	10 + 10	.73
Sum of standardized scores		-6 through +6		

## Procedure

At the first occasion, the APM was administered to the group of 51 participants as a whole and the procedure of the instructions for administration (Raven et al., 1998) was followed except for the following. After 20 minutes, the test administrator asked the participants to underscore on their response forms the item they were working on at that moment. This was repeated after 30 minutes (this time with a double underscore) and also after 40 minutes (triple underscore). The participants were allowed to work as long as they needed to try all 36 items, and they were allowed to leave the room when they finished the test.

At the second occasion, the APM was administered to the group of 51 plus 397 participants as a whole and the procedure of the instructions for administration of the timed version (Raven et al., 1998) was followed: The participants were informed that they were allowed to work for 20 minutes at the test. At the second occasion, all 51 plus 397 participants were also administered the six paper-and-pencil intelligence tests.

## Results

The correlation between the score of the untimed version at completion and the score of the 20-minute timed version of the APM was .75 (see Table 2). The correlation between the score after 20 minutes (untimed) and at completion (untimed) was .74. The correlation between the score after 20 minutes (untimed) and the score of the 20 minutes timed version of the APM was .69.

The mean scores after the successive intervals were increasing while the increases were negatively accelerated (see Table 3): The contribution to the score in the first 20 minutes was far greater than in the following 20 minutes or at completion of the test.

For the group of 51 students who did the APM twice, the difference between the mean score after 20 minutes on the untimed version ( $M = 20.51$ ,  $SD = 3.87$ ) and the mean score of the 20-minute timed version of the APM ( $M = 24.65$ ,  $SD = 3.30$ ) was statistically significant,  $t(50) = 10.27$ ,  $p < .0001$ ,  $d = 1.15$ . On the other hand, the mean score of the 51 students after 20 minutes on the untimed version of the APM

**Table 2**  
**Correlations Between the Untimed Version of the Raven Advanced Progressive Matrices Test (APM), the 20-Minute Timed Version, and the Six Tests**

	<i>N</i>	APM 30-Minute	APM 40-Minute	APM Completed	APM 20-Minute Timed	Six Tests Sum
APM 20-min.	51	.92	.87	.74	.69	.53
APM 30-min.	51		.95	.81	.71	.61
APM 40-min.	51			.86	.74	.60
APM completed	51				.75	.44 (.47)
APM 20-min. timed	51					.55
APM 20-min. timed	397					.42 (.23)

Note: *p* values are in parentheses.

**Table 3**  
**Descriptive Statistics of the Untimed Version of the Raven Advanced Progressive Matrices Test (APM), the 20-Minute Timed Version, and the Six Tests**

<i>N</i>	APM 20-Minute	APM 30-Minute	APM 40-Minute	APM Completed	APM 20-Minute Timed	Six Tests Sum
51	20.51 (3.87)	23.69 (3.63)	26.08 (3.94)	28.24 (3.86)	24.65 (3.30)	-.07 (3.96)
397					21.19 (4.29)	.01 (3.88)

Note: The table shows means with standard deviations in parentheses.

( $M = 20.51$ ,  $SD = 3.87$ ) did not differ from the mean score of the group of 397 students on the 20-minute timed version ( $M = 21.19$ ,  $SD = 4.29$ ):  $t(446) = -1.08$ ,  $p = .28$ ,  $d = 0.17$ .

The scores on the six paper-and-pencil tests were standardized and summed into one score. This score may be regarded as an intelligence measure. The correlation between the APM and other intelligence tests is reported to lie between .40 and .75 (Raven et al., 1998), whereas the correlation between the APM and different subtests of these intelligence tests is reported to lie between .24 and .60 (Raven et al., 1998). For present purposes, we consider an expected correlation of .45 to be reasonable, as the intelligence measure is based on just six subtests. We find that four correlations actually exceed this value, whereas two do not. Adopting a one-sided test ( $r < .45$ ), we thus need not consider the former four correlations. The latter two do not deviate from .45 ( $r = .44$ ,  $p = .47$  and  $r = .42$ ,  $p = .23$ ; see Table 2).

In the group of 51 students who completed the APM twice, the correlation between their 20-minute timed APM score and the score on the six tests was .55 (see Table 2). In the group of 397 students who only completed the timed version of the APM, this correlation was .42 (see Table 2). The statistical significance of the difference between these correlations was tested using Fisher's  $z_r$  transformation:  $z = 1.11$  (*ns*).

In the group of 51 students, the correlation between their 20-minute *untimed* APM score and the score on the six tests was .53 (see Table 2). The statistical significance of the difference between this correlation and the correlation between the 20-minute timed APM score and the score on the six tests in the group of 397 (.42) was also tested using Fisher's  $z_r$  transformation:  $z = 0.93$  (*ns*).

In the group of 51 students, the correlation between their *completed* APM score and the score on the six tests was .44 (see Table 2). The statistical significance of the difference between this correlation and the correlation between the 20-minute timed APM score and the score on the six tests in the group of 397 (.42) was also tested using Fisher's  $z_r$  transformation:  $z = 0.16$  (*ns*).

The mean scores on the 20-minute timed version of the APM of the two groups were 24.65 ( $N = 51$ ,  $SD = 3.30$ ) and 21.19 ( $N = 397$ ,  $SD = 4.29$ ) (see Table 3); this difference was statistically significant,  $t(446) = 5.54$ ,  $p < .001$ ,  $d = 0.90$ . The mean scores of the two groups on the six tests were  $-.07$  ( $N = 51$ ,  $SD = 3.96$ ) and  $.01$  ( $N = 397$ ,  $SD = 3.88$ ) (see Table 3), representing no statistically significant difference,  $t(446) = -.14$ , *ns*,  $d = 0.02$ .

## Discussion

In the untimed version of the APM, the score after 20 minutes is a reasonable predictor of the score after completion:  $r = .74$ . Of course, the scores after 30 and 40 minutes are increasingly better predictors.

The 51 students were as intelligent as the students in the group of 397 (based on the six tests sum), but they were approximately 3.5 points better at the 20-minute timed version of the APM. This we consider the effect of learning, because their scores on the 20-minute timed version of the APM were also better (approximately 4.6 points) than their scores after 20 minutes on the untimed version.

Could the 20-minute timed version of the APM also be considered a good predictor of the untimed APM score at completion? In the group of 51 participants, the correlation is .75. If the score on the six intelligence tests is taken as a criterion, the correlations of the two groups are not different: .55 ( $N = 51$ ) and .42 ( $N = 397$ ). For the group of 51 students, this is the correlation between the score of the six tests and the score of the *second* occasion for the APM. If the score of the *first* occasion is taken, the correlation was .53. Because the correlations between criterion and the APM score at the first and second occasions did not differ from each other, we are led to the conclusion that the 20-minute timed version of the APM can also be considered a good predictor of the untimed APM score at completion.

Could the timed version of the APM additionally involve a speed factor as well? Our results suggest this might not be the case. The correlation between the 20-minute timed version of the APM and the score on the six tests did not differ from the correlation between the untimed APM score at completion and the score on the six tests.

Bors and Stokes (1998) found a correlation of .88 between their short form and the APM, but this correlation reflects the relationship between a subset of the APM and the whole APM administered *in one session*. On the basis of this result, it cannot be decided how well their short form predicts scores on the whole APM, because the score on the short form depends on the achievement of the participants on the APM items “outside” the short form that they also tried.

Arthur and Day (1994) reported a correlation of .66 between their short form and the APM. They administered the short and long forms in separate sessions, like we did. We found correlations of .74 and .75 between the APM score and the scores after 20 minutes (untimed) and the 20-minute timed version, respectively. The results of our study allow for the conclusion that scores on the 20-minute timed version of the APM do predict very well scores on the untimed version, that is, the whole APM. As a consequence, it does not seem necessary to construe a short version of the APM by selecting items. In fact, keeping the APM as it is avoids possible loss of validity as a consequence of changing the task it represents. Although, for instance, it might be the case that the first nine items seem to have little to add to the discrimination between participants (cf. Arthur & Day, 1994; Bors & Stokes 1998), these items do contribute to the score of the participants in terms of *real* practice and experience that are indispensable ingredients of the APM as a power test (cf. Raven et al., 1993).

Administering the APM as a speed test, even with a limit of 20 minutes, is better than administering a selection of 12 items from the APM, even if this subset represents the APM quite well in a psychometric sense, because the selection also represents a quite different task for the participant, leaving the intermittent items out.

## References

- Arthur, W., Jr., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement, 54*, 394-403.
- Arthur, W., Jr., Tubre, T. C., Paul, D. S., & Sanchez-Ku, M. L. (1999). College-sample psychometric and normative data on a short form of the Raven Advanced Progressive Matrices Test. *Journal of Psychoeducational Assessment, 17*, 354-361.
- 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.
- 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.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.
- Elshout, J. J. (1976). *Karakteristieke moeilijkheden in het denken* [Characteristic difficulties in thinking]. Unpublished doctoral dissertation. Amsterdam: Universiteit van Amsterdam.



Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Raven, J., Raven, J. C., & Court, J. H. (1993). *Raven manual section 1: General overview*. Oxford: Oxford Psychologists Press.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual section 4: Advanced Progressive Matrices*. Oxford: Oxford Psychologists Press.

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