

# **Influential Article Review - Creation and validation of the Ravens Advanced Progressive Matrices**

**John Brock**

**Priscilla Lucas**

**Nora Stevens**

*This paper examines education. We present insights from a highly influential paper. Here are the highlights from this paper: The Ravens Advanced Progressive Matrices (APM) is a widely used measure of general intelligence (g), both across settings and cultures. Due to its lengthy 40-min administration time, several researchers have developed short-form scales, yet these forms typically yield a significantly lower reliability. This article describes the creation of an 18-item short form (APM-18) and its validation in three samples of Southwestern U.S. university students (total N = 633). The APM-18 shows similar psychometric properties to both the previously published 36-item long form and 12-item short form, but retains a reliability estimate closer to the original APM. This, plus the shorter administration time (25 min) relative to the complete APM (40-60 min), makes it useful for time-constrained or mass-testing situations. For our overseas readers, we then present the insights from this paper in Spanish, French, Portuguese, and German.*

*Keywords: general intelligence, Ravens Advanced Progressive Matrices, culture-fair tests, personality, timed tests*

## **SUMMARY**

- Each sample in Study 2 used different methods of assessing the convergent validity of the APM-18. Sample 1 focused on relationships between the APM-18 and other standard measures of intelligence and academic achievement; whereas Sample 2 examined the relationship between the APM-18, creativity, self-reported ACT scores, and Big Five personality traits. Both studies confirmed that the APM-18 is related to these measures in a predictable manner.
- Generally speaking, both the APM-18 and the embedded APM-12 showed the same pattern of correlations with the other measures used in these studies. However, the higher IC of the APM-18 suggests that it may be better at detecting individual variation in *g*.
- Each of the 18 items used in this new APM-18 test was chosen to maintain the progressive difficulty of both the long form and the short form. Unsurprisingly, although the APM-18's reliability was lower than that of the APM-36, it was higher than that of the APM-12 developed by Arthur and Day. Furthermore, the patterns of correlation with other measures of intelligence are virtually identical to the APM-12, which has, in previous studies, been shown to mimic the APM-36 results.

- Combined with an average administration time of 17.53 min, these findings suggest that the APM-18 may work well as a compromise for researchers who want a quite accurate measure of general intelligence in a quite short amount of time.
- The cross-validation in the three samples reported here is an initial attempt to collect normative data for the APM-18. Our results may generalize only to other college students. However, the APM-18's short administration time, high IC, reasonable validity, and ease of administration by paper and pencil in large college classroom settings make it ideal for behavioral science studies where researchers want a reasonably fast, accurate intelligence score as part of a larger questionnaire battery.

## HIGHLY INFLUENTIAL ARTICLE

We used the following article as a basis of our evaluation:

Sefcek, J. A., Miller, G. F., & Figueredo, A. J. (2016). Development and Validation of an 18-Item Medium Form of the Ravens Advanced Progressive Matrices. *SAGE Open*. <https://doi.org/10.1177/2158244016651915>

This is the link to the publisher's website:

<https://journals.sagepub.com/doi/full/10.1177/2158244016651915#articleCitationDownloadContainer>

## INTRODUCTION

The Ravens Progressive Matrices Test, developed by Raven (1941) as a measure of general intelligence (*g*), has undergone many revisions, ranging from colored versions for children to the standard and advanced matrices for adults of different cognitive levels. The most recent published version is the Ravens Advanced Progressive Matrices (APM; Raven, Raven, & Court, 1993), which was developed for higher ability adult populations (i.e., college-level and above). This test is constructed of 36 items of increasing difficulty broken into three 12-item sets; in each item, the examinee is asked to complete a visual pattern by choosing one of eight possible solutions.

Due to its nonverbal format, the APM is purported to be a culturally fair, unbiased measure of fluid intelligence (Cattell, 1963), educative ability (J. Raven et al., 1993), or, as we will refer to it, general intelligence (*g*; Spearman, 1927), and has shown itself to be especially useful in situations where English is not an individual's primary language. As such, the Standard and Advanced Progressive Matrices have been used extensively in many applied settings in the United States (e.g., Ackerman, 1992) and across many cultures (Owen, 1992; J. C. Raven, 2000; Rushton, Cvorovic, & Bons, 2007). However, the positive aspects of this test are marred by its lengthy administration time (40-60 min), making it difficult to use in time-constrained multivariate research or classroom settings.

In answer to these various limitations, Arthur and Day (1994) developed a 12-item short form of the APM (which we call APM-12), with an administration time of 15 min. Several studies have shown that this 12-item form shows acceptable psychometric properties (e.g., Cronbach's alpha, test-retest reliability, convergent validity; see Arthur, Tubre, Paul, & Sanchez-Ku, 1999, for review). However, this short form shows relatively low and variable internal consistency (IC). For example, Cronbach's alphas range from .58 to .66 for short form itself to and .72 to .73 for the 12 short-form items extracted from the full 36-item version (Arthur & Day, 1994).

More recently, Hamel and Schmittmann (2006) have argued that the complete 36-item APM can be administered as a 20-min speed test. Scores on this speeded form of the APM show strong correlations with scores on slower timed (40 min,  $r = .74$ ) and untimed versions ( $r = .75$ ) of the APM. However, these authors failed to report the IC of the Speed Test Scale. We also suspect that giving typical adults only 20 min to complete 36 very challenging abstract reasoning problems might impose undue stress.

The purpose of the current study was to develop a medium-form version of the APM that resulted in higher IC than the 12-item version (APM-12), but shorter administration time than the full 36-item APM (APM-36)—a combination of features that might be useful for time-constrained and mass-testing situations. Here, we report the development and construct validity of this 18-item scale.

## **STUDY 1: SCALE CONSTRUCTION AND CONSTRUCT VALIDITY**

### **Method**

#### ***Participants***

A total of 633 students (198 male, 435 female) from three southwestern universities participated in this study as a partial requirement for experimental course credit. The mean age for participants was 20.92,  $SD = 4.07$  ( $M_{male} = 20.85$ ,  $SD_{male} = 3.90$ ;  $M_{female} = 20.96$ ,  $SD_{female} = 4.15$ ). Ages ranged from 17 to 58 years old (male = 18-41, female = 17-58).

### **Measures**

#### ***The Ravens Advanced Progressive Matrices 18-Item Short Form (APM-18)***

This 18-item short-form version of the APM is printed in a booklet format on 8½" by 11" white paper, with each test item printed on a separate page. The first four pages of the test booklet contain three example items (Practice Items 1, 5, and 9 from APM-36) to explain the task.

The 18 actual test items were derived by adding six items from the longer 36-item version (J. Raven et al., 1993) to Arthur and Day's (1994) published 12-item version. Arthur and Day used Items 1, 4, 8, 11, 15, 18, 21, 23, 25, 30, 31, and 35 from the 36-item APM based on a set of three decision rules, which can be summed up as (a) dividing the APM into 12, three-item sections based on difficulty; (b) taking the item with the highest item-total correlation for each section; and (c) in the case of a tie, including the item that resulted in the largest drop in IC if it was excluded from the full test. Following these same rules, we added six more items of increasing difficulty—two that were easy (96% and 75% of examinees from the normative sample answered correctly), two that were moderate (50% and 48% of examinees from the normative sample answered correctly), and two that were difficult (37% and 32% of examinees from the normative sample answered correctly). These items (2, 20, 22, 24, 34, and 32) were integrated of difficulty to mimic their presentation order in the original APM.

#### ***Procedure***

The new APM-18 test was given in classroom settings with several examinees at a time. This was done because this test was developed as a measure of  $g$  that could be used in environments such as classrooms, where there are time limits on research sessions. In one subsample ( $n = 175$ ), tests were given with no time constraints, but with completion times recorded, to determine the average time needed for completion. The other two subsamples ( $n = 232$  and  $n = 226$ ) were constrained to finish the test within 25 min, with no individual completion times recorded.

#### ***Analyses***

All statistical analyses were conducted using SAS Version 8.2 (SAS Institute, 1999). Cronbach's alphas and bivariate correlations were computed using the PROC CORR procedure. Tests for mean differences between sexes were calculated through  $t$  test (PROC TTEST) procedures. Hierarchical general linear models (GLMs) were tested using PROC GLM.

#### ***Results***

IC estimates were computed by using Cronbach's alpha. The IC of the APM-18 scale yielded moderate reliability ( $\alpha = .79$ ). This alpha is lower than normative IC reports for the APM-36 ( $\alpha = .84$ ; Forbes, 1964), but higher than those for the APM-12 (ranging from  $\alpha = .58$ -.66; see Arthur et al., 1999). Furthermore, the alpha of the APM-18 was larger than that of the embedded APM-12 ( $\alpha = .73$ ). Table 1 shows the results for each of the APM-18 items, with respect to their item-total correlations, item difficulties, and scale  $\alpha$  of the

overall scale if the item is deleted. As seen, deleting any item reduces the overall reliability of the scale, suggesting that all items should be retained.

The mean APM-18 score was 9.73,  $SD = 3.59$  ( $M_{\text{male}} = 10.43$ ,  $SD_{\text{male}} = 3.52$ ;  $M_{\text{female}} = 9.41$ ,  $SD_{\text{female}} = 3.59$ ), with a range of 18. For the subsample in which completion times were recorded ( $n = 175$ ), the mean test completion time was 17.5 min ( $SD = 4.67$ ), with a range of 7 to 25 min; 21% of the participants took longer than 20 min, but no one took longer than 25 min. In this subsample, there was a significant positive relationship between the amount of time it took for participants to take the test and their APM-18 score ( $r = .41$ ,  $p < .001$ ), but there was no relationship between age and APM-18 score, or age and time required to complete the test (for each  $r \leq .03$ ,  $p \geq .71$ ). However, for the complete sample ( $N = 633$ ), younger participants scored a little higher (age and APM-18 scores correlated  $r = -.15$ ,  $p < .001$ ), and males scored a little higher—sex (female = 0, male = 1) and APM-18 scores correlated  $r = .13$ ,  $p < .001$ .

Hierarchical GLMs were tested to explore whether the apparent differences in male and female APM-18 scores might have been indirectly attributable to the relationship between age and APM-18 scores. This model defined the APM-18 score as the criterion variable, with the ordered predictor variables being age and then sex. The hierarchical model was designed to allow age to absorb as much variance as possible, with sex entered into the model only afterward. Using this model, both GLMs indicated a significant effect for age ( $F = 15.39$ ,  $p < .001$ ) and then also for sex after age had been statistically controlled ( $F = 10.66$ ,  $p = .001$ ).

## **Discussion**

The results presented here suggest that the APM-18 may serve as a useful compromise between the lower reliability APM-12 and the much longer APM-36. The hierarchical GLMs identify both age and sex to be significant predictors of APM-18 scores, with younger individuals and males generally scoring higher. These results are consistent with many previous studies looking at general intelligence (e.g., Jackson & Rushton, 2006). Results of Study 1, however, do not test the convergent validity of this scale relative to other measures of intelligence. Study 2 was designed to do this.

## **STUDY 2: CONVERGENT VALIDITY**

Study 2 was conducted to assess the convergent validity of the APM-18 with other measures of intelligence, academic achievement, and personality. To do so, we tested two separate subsamples ( $n = 193$  and 229) taken from the Study 1, each of which used different criterion measures. In Sample 1, two widely used measures of adult intelligence were used: the Mill-Hill Vocabulary Scale-Multiple Choice Sets A & B (MHV-MC; J. Raven, Raven, & Court, 1997), developed to be used in conjunction with the APM-36 as a measure of reproductive ability, that is, the ability to store and retrieve information (J. C. Raven, 1989); and the Shipley Institute of Living Scale (SILS; Zachary, 1986), which is a stand-alone intelligence test comprised of two subscales—Vocabulary, which tests crystallized intelligence, and Abstraction, which tests fluid intelligence. Also, we examined academic performance via self-reported grade point average (GPA) and scholastic aptitude test (SAT) scores. In Sample 2, we examined correlations between APM-18 scores and Big Five personality dimensions assessed with the NEO Five-Factor Inventory (NEO-FFI) Scale (Costa & McCrae, 1992), and verbal and drawing creativity (Miller & Tal, 2007). In addition, ACT scores were collected in this second sample.

## **Method**

### ***Participants***

Sample 1 was comprised of 193 students (94 male, 99 female) from an introductory psychology course at the University of Arizona. Mean age of participants was 19.11,  $SD = 1.62$  ( $M_{\text{male}} = 19.23$ ,  $SD_{\text{male}} = 1.07$ ;  $M_{\text{female}} = 19.01$ ,  $SD_{\text{female}} = 2.00$ ). Due to the length of time required to administer the APM-18, the Shipley, and the Mill–Hill, 10 participants did not complete the Mill–Hill Test. We urged participants to record their SAT, ACT, and GPA scores only if they were certain of them; due to this constraint, many of these scores were also missing.

Sample 2 was comprised of 229 students (65 male, 164 female) from various undergraduate courses at the University of New Mexico. Mean age of participants was 20.19,  $SD = 3.43$  ( $M_{\text{male}} = 21.05$ ,  $SD_{\text{male}} = 5.01$ ;  $M_{\text{female}} = 19.85$ ,  $SD_{\text{female}} = 2.48$ ). Again, we urged participants to record their ACT scores only if they were certain of them, leaving us with ACT scores for only 129 participants.

## **Measures**

### ***APM-18***

The APM-18 consisted of the same items identified in Study 1. In Sample 1, the form was presented first in a series of measures examining adult intelligence. In Sample 2, it was presented in the middle of a questionnaire packet concerning personality, creativity, sexual behavior, and intelligence.

### ***The SILS***

The SILS (Zachary, 1986) is a timed (10 min per subscale), 60-item self-report measure that examines both verbal intelligence (40 items) and abstract intelligence (20 items). The test is considered appropriate for average English-speaking individuals from 14 to adult ages, who are motivated test takers. Validities and norms published in the manual were taken from a sample of 322 army recruits. Split-half reliabilities for each subscale are reported as .87 for Vocabulary, .89 for Abstraction, and .92 for the total score.

### ***The MHV-MC***

The MHV-MC (J. Raven et al., 1997) is a 68-item self-administered multiple-choice vocabulary test designed to complement the APM-36. Whereas the APM aimed to measure an individual's ability to solve novel problems and think in novel ways (i.e., fluid intelligence), the Mill–Hill aimed to measure an individual's ability to recall learned information (i.e., crystallized intelligence). To this extent, it indicates educational attainments, cultural background, and familiarity with the test's language. The Mill–Hill typically shows split-half reliabilities more than .90 and test–retest reliabilities ranging between .87 and .95 (Raven et al., 1997).

### ***Academic performance***

Academic performance was measured by self-reported GPAs and SAT scores in Sample 1. Sample 2 participants were asked for SAT and ACT scores. A variety of studies have identified moderate to strong correlations between these academic achievement and aptitude measures, and a variety of other traits, including intelligence, personality, and psychopathology (Barton, Dielman, & Cattell, 1971; Brown, 1994; Dyer, 1987; Mouw & Khanna, 1993).

### ***NEO-FFI***

The NEO–FFI (Costa & McCrae, 1992) is the most widely used measure in research on the Five–Factor model of personality. It is a shortened version of the 240–item Revised NEO Personality Inventory (NEO–PI–R; Costa & McCrae, 1992), comprised of 60 items that measure five global personality factors (12 items per factor): Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. In our version, participants rated degree of agreement with statements about their personalities and behavioral propensities on a 5–point scale ranging from –2 (strongly disagree) to 0 (neutral) to +2 (strongly agree). This scale has shown strong

IC, with Cronbach's alphas ranging between .74 to .89, for each factor, and consistent cross-cultural validity (McCrae & Costa, 1997).

### *Verbal and drawing creativity tasks*

Participants completed six 2-min verbal creativity tasks and eight 1-min drawing creativity tasks (Miller & Tal, 2007). Because a mating-oriented mind-set promotes creativity (Griskevicius, Cialdini, & Kenrick, 2006), participants were asked to complete these tasks as creatively as possible with the intention of attracting a romantic partner. Examples of verbal tasks included writing answers to thought-provoking questions, such as "How would you keep a marriage exciting after the first couple of years?" "What do you hope the world will be like in 100 years?" and "Imagine that all clouds had really long strings hanging from them—strings hundreds of feet long. What would be the implications of that fact for nature and society?" There were two types of drawing tasks, four abstract (e.g., "Please draw an abstract symbol, pattern, or composition that represents your happiness as a child doing a favorite activity") and four representational (e.g., "In the space below, please draw an animal that you admire for its strength, grace, speed, or beauty"). Each participant's responses to each of the 14 creativity tasks were scored independently by four raters on a 1- to 5-point creativity scale. The resulting composite verbal creativity and drawing creativity measures showed high interrater reliability and IC (Cronbach's alphas = .91 and .90, respectively; Miller & Tal, 2007).

## RESULTS

### Sample 1

IC estimates were computed using Cronbach's alpha. The APM-18 showed moderate IC ( $\alpha = .71$ ), with the embedded APM-12 yielding a slightly lower value ( $\alpha = .63$ ). Although these internal consistencies are lower than those reported in Study 1, they are still moderate in strength.

The mean APM-18 score was 10.68,  $SD = 3.25$  ( $M_{\text{male}} = 11.07$ ,  $SD_{\text{male}} = 3.13$ ;  $M_{\text{female}} = 10.31$ ,  $SD_{\text{female}} = 3.34$ ), with a range of 13 (four to 17). There was no relationship between APM-18 score and age ( $r = .03$ ,  $p = .71$ ), or sex ( $r = .11$ ,  $p = .10$ ). Mean scores for each of the measures of intelligence can be seen in Table 2. Due to the significant sex difference between APM-18 scores in Study 1, mean sex differences on all measures in this study were checked via *t* tests. There were no significant sex differences for any of the intelligence measures in this sample, except for a moderate male advantage on self-reported SAT scores ( $t = -3.00$ ,  $p = .003$ ). Therefore, the remaining analyses were conducted on the full sample rather than by sex.

As seen in Table 3, both the APM-18 and embedded APM-12 correlated significantly with most of the other measures of intelligence and academic achievement and aptitude used in this sample. Specifically, both the APM scales correlated positively and significantly most strongly with the Shipley Abstraction scale and self-report SAT scores. This is not surprising as the APM is designed to be a measure of *g*, which may be most easily identified in relation to abstract, analytical measures, of which the Shipley abstraction is one, and the SAT contains an analytical subscale.

### Sample 2

IC estimates were again computed using Cronbach's alpha. As in Study 1 and Sample 1 of this study, the APM-18 showed moderate reliability ( $\alpha = .79$ ), whereas the embedded APM-12 again shows slightly lower reliability ( $\alpha = .74$ ). The mean APM-18 score was 9.53,  $SD = 3.57$  ( $M_{\text{male}} = 10.29$ ,  $SD_{\text{male}} = 3.98$ ;  $M_{\text{female}} = 9.23$ ,  $SD_{\text{female}} = 3.36$ ), with a range of 1 to 18. There was no relationship between APM-18 score and age ( $r = -.03$ ,  $p = .61$ ), but there was between APM-

18 score and sex ( $r = .13$ ,  $p = .04$ ), with males again scoring slightly higher ( $t = -2.04$ ,  $p < .05$ ). There were no other sex differences on the other intelligence measures (for all  $t_s \geq -1.20$ ,  $p > .05$ ). Table 4 shows mean scores for the intelligence measures and NEO-FFI factors.

As seen in Table 5, both the APM-18 and embedded APM-12 scores were significantly positively related to verbal creativity ( $r_s = .36$  and  $.32$ ,  $p < .001$ , respectively), drawing creativity ( $r_s = .34$  and  $.29$ ,  $p < .001$ , respectively), and self-reported ACT score ( $r_s = .44$  and  $.45$ ,  $p < .001$ , respectively). In addition, the APM scales were positively related to Openness ( $r_s = .26$  and  $.25$ ,  $p < .01$ , respectively) and negatively to Conscientiousness (for each,  $r = -.16$ ,  $p < .05$ ), a finding consistent with previous research (Ackerman & Heggestad, 1997; Moutafi, Furnham, & Crump, 2003).

## DISCUSSION

Each sample in Study 2 used different methods of assessing the convergent validity of the APM-18. Sample 1 focused on relationships between the APM-18 and other standard measures of intelligence and academic achievement (e.g., verbal intelligence tests, self-reported GPA, and SAT scores); whereas Sample 2 examined the relationship between the APM-18, creativity, self-reported ACT scores, and Big Five personality traits. Both studies confirmed that the APM-18 is related to these measures in a predictable manner. Generally speaking, both the APM-18 and the embedded APM-12 showed the same pattern of correlations with the other measures used in these studies. However, the higher IC of the APM-18 suggests that it may be better at detecting individual variation in  $g$ .

## CONCLUSION

Each of the 18 items used in this new APM-18 test was chosen to maintain the progressive difficulty of both the long form (APM-36) and the short form (APM-12). Unsurprisingly, although the APM-18's reliability was lower than that of the APM-36, it was higher than that of the APM-12 developed by Arthur and Day (1994). Furthermore, the patterns of correlation with other measures of intelligence are virtually identical to the APM-12, which has, in previous studies, been shown to mimic the APM-36 results (Arthur & Day, 1994; Arthur et al., 1999). Combined with an average administration time of 17.53 min (25 min maximum), these findings suggest that the APM-18 may work well as a compromise for researchers who want a quite accurate measure of general intelligence in a quite short amount of time. The cross-validation in the three samples reported here is an initial attempt to collect normative data for the APM-18. Our results may generalize only to other college students. However, the APM-18's short administration time, high IC, reasonable validity, and ease of administration by paper and pencil in large college classroom settings make it ideal for behavioral science studies where researchers want a reasonably fast, accurate intelligence score as part of a larger questionnaire battery.

## APPENDIX

**TABLE 1**  
**ITEM-TOTAL CORRELATIONS AND ITEM DIFFICULTY FOR THE APM-18.**

	Item-total correlation	% correct	Scale $\alpha$ if deleted
1	.3228	.9289	.7798
2	.3871	.9021	.7754
3	.4160	.8483	.7734
4	.4957	.8168	.7678
5	.3896	.8310	.7752
6	.3973	.7899	.7747
7	.3599	.5861	.7772
8	.3975	.6019	.7747
9	.4800	.5071	.7689
10	.4273	.4801	.7726
11	.2512	.3612	.7846
12	.3214	.3207	.7799
13	.4188	.4739	.7732
14	.3480	.3175	.7781
15	.3211	.2401	.7799
16	.3357	.3144	.7789
17	.2123	.2101	.7872
18	.2734	.1974	.7831

Note. APM-18 = Ravens Advanced Progressive Matrices 18-Item Short Form.

**TABLE 2**  
**DESCRIPTIVE STATISTICS FOR MEASURES OF INTELLIGENCE IN SAMPLE 1 (N = 193).**



	Total		Male		Female	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
APM-18	10.68	3.25	11.07	3.13	10.31	3.34
Shipley	43.47	6.80	42.96	7.69	43.96	5.82
Shipley Verbal	27.65	5.77	27.29	6.65	27.99	4.80
Shipley Abstraction	31.65	5.04	31.34	5.47	31.94	4.59
Mill-Hill	46.87	6.91	46.51	7.98	47.24	5.68
GPA	3.10	0.59	3.03	0.62	3.17	0.56
SAT	1,140.98 <sup>b</sup>	163.66	1,179.32	143.49	1,099.90	174.66

Note. APM-18 = Ravens Advanced Progressive Matrices 18-Item Short Form; GPA = grade point average; SAT = scholastic aptitude test.  
<sup>b</sup>Indicates mean differences as a function of sex.

**TABLE 3**  
**CORRELATIONS AMONG APM-18, APM-12, AND OTHER INTELLIGENCE AND**  
**ACADEMIC ACHIEVEMENT MEASURES IN SAMPLE 1.**

	Mill-Hill (n = 183)	Shipley Abstraction (n = 193)	Shipley Vocabulary (n = 193)	GPA (n = 175)	SAT (n = 144)
APM-18	.22*	.49*	.12	.17*	.34*
APM-12 (embedded)	.26*	.47*	.16*	.12	.32*

Note. APM-18 = Ravens Advanced Progressive Matrices 18-Item Short Form; APM-12 = Ravens Advanced Progressive Matrices 12-Item Short Form; GPA = grade point average; SAT = scholastic aptitude test.

\*p < .05.

**TABLE 4**  
**DESCRIPTIVE STATISTICS FOR MEASURES OF INTELLIGENCE, ACADEMIC**  
**ACHIEVEMENT, AND PERSONALITY IN SAMPLE 2 (N = 229).**

	Total		Male		Female	
	M	SD	M	SD	M	SD
APM-18	9.53 <sup>†</sup>	3.57	10.29	3.98	9.23	3.36
Verbal Creativity	2.57	0.46	2.56	0.60	2.57	0.40
Drawing Creativity	2.62	0.68	2.60	0.76	2.63	0.65
ACT (n = 129)	23.16	4.20	23.34	4.21	23.08	4.22
Openness	13.83	10.69	13.94	9.84	13.78	11.03
Conscientiousness	11.44	10.78	10.17	11.00	11.93	10.68
Extraversion	12.68	10.69	12.11	11.22	12.91	10.49
Agreeableness	4.74	11.08	3.77	11.27	5.13	11.01
Neuroticism	2.49	12.02	-1.41	12.20	4.04	11.62

Note. APM-18 = Ravens Advanced Progressive Matrices 18-Item Short Form.

<sup>†</sup>Indicates mean differences as a function of sex.

**TABLE 5**  
**CORRELATIONS AMONG APM-18, APM-12, AND OTHER INTELLIGENCE, ACADEMIC**  
**ACHIEVEMENT, AND PERSONALITY MEASURES IN SAMPLE 2.**

	Verbal Creativity (n = 225)	Drawing Creativity (n = 225)	ACT (n = 129)	O (n = 224)	C (n = 226)	E (n = 221)	A (n = 223)	N (n = 225)
APM-18	.35*	.29*	.44*	.26*	-.16*	-.03	-.11	.02
APM-12 (embedded)	.31*	.24*	.45*	.25*	-.16*	-.03	-.07	.00

Note. O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; N = Neuroticism; APM-18 = Ravens Advanced Progressive Matrices 18-Item Short Form; APM-12 = Ravens Advanced Progressive Matrices 12-Item Short Form.

\*p < .05.

## REFERENCES

- Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, 77, 598-614.
- Ackerman, P. L., Heggstad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, 121, 219-245.
- Arthur, W., 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., 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.
- Barton, K., Dielman, T. E., Cattell, R. B. (1971). The prediction of school grades from personality and IQ measures. *Personality*, 2, 325-333.
- Brown, M. W. (1994). Cognitive, interest, and personality variables predicting first-semester GPA. *Psychological Reports*, 74, 605-606.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- Costa, P. T., McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources.
- Dyer, E. D. (1987). Can university success and first-year job performance be predicted from academic achievement, vocational interest, personality and biographical measures? *Psychological Reports*, 61, 655-671.
- Forbes, A. R. (1964). An item analysis of the advanced matrices. *British Journal of Educational Psychology*, 34, 223-236.
- Griskevicius, V., Cialdini, R. B., Kenrick, D. T. (2006). Peacocks, Picasso, and parental investment: The effects of romantic motives on creativity. *Journal of Personality and Social Psychology*, 91, 52-66.
- Hamel, R., Schmittmann, V. D. (2006). The 20-minute version as a predictor of the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, 66, 1039-1046.
- Jackson, D. N., Rushton, J. P. (2006). Males have greater g: Sex differences in general mental ability from 100,000 17- to 18-year-olds on the Scholastic Assessment Test. *Intelligence*, 34, 479-486.
- McCrae, R. R., Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509-516.
- Miller, G. F., Tal, I. R. (2007). Schizotypy versus openness and intelligence as predictors of creativity. *Schizophrenia Research*, 93, 317-324.
- Moutafi, J., Furnham, A., Crump, J. (2003). Demographic and personality predictors of intelligence: A study using the NEO personality inventory and the Myers-Briggs type indicator. *European Journal of Personality*, 17, 79-94.
- Mouw, J. T., Khanna, R. K. (1993). Prediction of academic success: A review of the literature and some recommendations. *College Student Journal*, 27, 328-336.
- Owen, K. (1992). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences*, 13, 149-159.
- Raven, J., Raven, J. C., Court, J. H. (1993). Raven manual: Section 1. General overview. Oxford, UK: Oxford Psychologists Press.
- Raven, J., Raven, J. C., Court, J. H. (1997). Mill Hill Vocabulary Scale: 1998 edition. Oxford, UK: Oxford Psychologists Press.
- Raven, J. C. (1941). Standardisation of progressive matrices. *British Journal of Medical Psychology*, 19, 137-150.
- Raven, J. C. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement*, 26, 1-16.

- Raven, J. C. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, 41, 1-48.
- Rushton, J. P., Cvorovic, J., Bons, T. A. (2007). General mental ability in South Asians: Data from three Roma (Gypsy) communities in Serbia. *Intelligence*, 35, 1-12.
- SAS Institute . (1999). SAS/STAT user's guide, version 8, volumes 1, 2, and 3. Cary, NC: Author.
- Spearman, C. (1927). *The abilities of man*. London, England: Macmillan.
- Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised Manual*. Los Angeles: Western Psychological Services

## **TRANSLATED VERSION: SPANISH**

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

## **VERSION TRADUCIDA: ESPAÑOL**

A continuación se muestra una traducción aproximada de las ideas presentadas anteriormente. Esto se hizo para dar una comprensión general de las ideas presentadas en el documento. Por favor, disculpe cualquier error gramatical y no responsabilite a los autores originales de estos errores.

## **INTRODUCCIÓN**

La Prueba de Matrices Progresivas Ravens, desarrollada por Raven (1941) como una medida de inteligencia general (g), ha sido sometida a muchas revisiones, que van desde versiones de colores para niños hasta matrices estándar y avanzadas para adultos de diferentes niveles cognitivos. La versión publicada más reciente es Ravens Advanced Progressive Matrices (APM; Raven, Raven, & Court, 1993), que fue desarrollado para poblaciones adultas de mayor capacidad (es decir, nivel universitario y superior). Esta prueba está construida de 36 elementos de dificultad creciente divididos en tres conjuntos de 12 elementos; en cada elemento, se pide al examinado que complete un patrón visual eligiendo una de las ocho soluciones posibles.

Debido a su formato no verbal, se pretendía que el APM es una medida culturalmente justa e imparcial de inteligencia fluida (Cattell, 1963), capacidad educativa (J. Raven et al., 1993), o, como nos referiremos a ella, inteligencia general (g; Spearman, 1927), y ha demostrado ser especialmente útil en situaciones donde el inglés no es el idioma principal de un individuo. Como tal, las Matrices Progresivas Estándar y Avanzada se han utilizado ampliamente en muchos entornos aplicados en los Estados Unidos (por ejemplo, Ackerman, 1992) y en muchas culturas (Owen, 1992; J. C. Raven, 2000; Rushton, Cvorovic, & Bons, 2007). Sin embargo, los aspectos positivos de esta prueba se ven empañados por su largo tiempo de administración (40-60 min), lo que dificulta su uso en la investigación multivariada limitada por el tiempo o en entornos de clase.

En respuesta a estas diversas limitaciones, Arthur y Day (1994) desarrollaron una forma corta de 12 elementos del APM (que llamamos APM-12), con un tiempo de administración de 15 min. Varios estudios han demostrado que este formulario de 12 elementos muestra propiedades psicométricas aceptables (por ejemplo, alfa de Cronbach, prueba-retest fiabilidad, validez convergente; véase Arthur, Tubre, Paul, & Sanchez-Ku, 1999, para su revisión). Sin embargo, esta forma abreviada muestra consistencia interna (IC) relativamente baja y variable. Por ejemplo, los alfas de Cronbach van de .58 a .66 para el propio formulario corto a y .72 a .73 para los 12 elementos de forma corta extraídos de la versión completa de 36 elementos (Arthur & Day, 1994).

Más recientemente, Hamel y Schmittmann (2006) han argumentado que el APM completo de 36 elementos puede administrarse como una prueba de velocidad de 20 minutos. Las puntuaciones en esta forma acelerada del APM muestran correlaciones fuertes con puntuaciones en versiones cronométricas más

lentas (40 min,  $r .74$ ) y versiones no cronometradas ( $r .75$ ) del APM. Sin embargo, estos autores no informaron el IC de la escala de prueba de velocidad. También sospechamos que dar a los adultos típicos sólo 20 minutos para completar 36 problemas de razonamiento abstracto muy desafiantes podría imponer estrés indebido.

El propósito del estudio actual era desarrollar una versión de forma media del APM que resultara en un IC más alto que la versión de 12 elementos (APM-12), pero un tiempo de administración más corto que el APM completo de 36 elementos (APM-36), una combinación de características que podrían ser útiles para situaciones de pruebas en masa y con limitaciones de tiempo. Aquí, informamos del desarrollo y la validez de la construcción de esta escala de 18 elementos.

## **ESTUDIO 1: CONSTRUCCIÓN A ESCALA Y VALIDEZ DE CONSTRUCCIÓN**

### **Método**

#### ***Participantes***

Un total de 633 estudiantes (198 hombres, 435 mujeres) de tres universidades del suroeste participaron en este estudio como un requisito parcial para el crédito de cursos experimentales. La edad media para los participantes fue de 20,92, SD a 4,07 (Mmale, 20,85, SDmale, 3,90; Mfemale 20,96, SDfemale 4,15). Las edades oscilaron entre los 17 y los 58 años (masculinos de 18 a 41 años, mujeres de 17 a 58 años).

### **Medidas**

#### ***El Ravens Advanced Progressive Matrices 18-Item Short Form (APM-18)***

Esta versión de formato corto de 18 artículos del APM se imprime en un formato de folleto en papel blanco de 81,2" por 11", con cada elemento de prueba impreso en una página separada. Las primeras cuatro páginas del folleto de prueba contienen tres elementos de ejemplo (elementos de práctica 1, 5 y 9 de APM-36) para explicar la tarea.

Los 18 elementos de prueba reales se derivaron añadiendo seis elementos de la versión más larga de 36 elementos (J. Raven et al., 1993) a la versión publicada de 12 artículos de Arthur y Day (1994). Arthur y Day usaron los puntos 1, 4, 8, 11, 15, 18, 21, 23, 25, 30, 31 y 35 de la APM de 36 puntos sobre la base de un conjunto de tres reglas de decisión, que se pueden resumir como (a) dividir el APM en 12, secciones de tres elementos basadas en dificultad; (b) tomar el elemento con la correlación elemento-total más alta para cada sección; y c) en el caso de un empate, incluido el elemento que resultó en la mayor caída de IC si se excluyó de la prueba completa. Siguiendo estas mismas reglas, añadimos seis elementos más de dificultad creciente: dos que eran fáciles (el 96% y el 75% de los examinadores de la muestra normativa respondieron correctamente), dos que eran moderados (50% y 48% de los examinadores de la muestra normativa respondieron correctamente), y dos que fueron difíciles (37% y 32% de los examinados de la muestra normativa respondieron correctamente). Estos elementos (2, 20, 22, 24, 34 y 32) se integraron de dificultad para imitar su orden de presentación en el APM original.

### ***Procedimiento***

La nueva prueba APM-18 se dio en el aula con varios exámenes a la vez. Esto se hizo porque esta prueba se desarrolló como una medida de  $g$  que podría ser utilizada en entornos como las aulas, donde hay límites de tiempo en las sesiones de investigación. En una submuestra ( $n = 175$ ), se realizaron pruebas sin restricciones de tiempo, pero con los tiempos de finalización registrados, para determinar el tiempo medio necesario para la finalización. Las otras dos submuestras ( $n = 232$  y  $n = 226$ ) se vieron restringidas para finalizar la prueba en un plazo de 25 min, sin que se registraran tiempos de finalización individuales.

### ***Análisis***

Todos los análisis estadísticos se realizaron utilizando SAS Versión 8.2 (Instituto SAS, 1999). Las correlaciones alfas y bivariadas de Cronbach se calcularon utilizando el procedimiento PROC CORR. Las

pruebas de diferencias medias entre sexos se calcularon mediante procedimientos de prueba t (PROC TTEST). Los modelos lineales generales jerárquicos (GMM) se probaron utilizando PROC GLM.

## Resultados

Las estimaciones de IC se calcularon utilizando el alfa de Cronbach. El IC de la escala APM-18 produjo una fiabilidad moderada (n.o 79). Este alfa es inferior a los informes de IC normativo para el APM-36 (a .84; Forbes, 1964), pero más alto que los del APM-12 (que oscila entre .58-.66; véase Arthur et al., 1999). Además, el alfa del APM-18 era mayor que el del APM-12 incrustado (a .73). La Tabla 1 muestra los resultados de cada uno de los elementos APM-18, con respecto a sus correlaciones de elemento total, dificultades del elemento y escala de la escala global si se elimina el elemento. Como se ha visto, la eliminación de cualquier elemento reduce la fiabilidad general de la escala, lo que sugiere que se deben conservar todos los elementos.

La puntuación media de APM-18 fue de 9,73, SD a 3,59 (Mmachos a 10,43, SD machos a 3,52; Mhembra 9,41, SDhembra 3,59), con un rango de 18. Para la submuestra en la que se registraron los tiempos de finalización (n a 175), el tiempo medio de finalización de la prueba fue de 17,5 min (SD a 4,67), con un rango de 7 a 25 min; 21% de los participantes tardaron más de 20 minutos, pero nadie tomó más de 25 minutos. En esta submuestra, hubo una relación positiva significativa entre la cantidad de tiempo que tardaron los participantes en tomar la prueba y su puntuación APM-18 ( $r = .41, p < .001$ ), pero no hubo relación entre la edad y la puntuación de APM-18, o la edad y el tiempo requeridos para completar la prueba (para cada  $r = .03, p = .71$ ). Sin embargo, para la muestra completa (N = 633), los participantes más jóvenes obtuvieron una puntuación un poco más alta (edad y puntuaciones de APM-18 correlacionadas  $r = .15, p < .001$ ), y los machos obtuvieron una puntuación un poco más alta: sexo (femenino 0, masculino 1) y puntuaciones APM-18 correlacionadas  $r = .13, p < .001$ .

Se probaron GLM jerárquicos para explorar si las diferencias aparentes en las puntuaciones APM-18 masculinas y femeninas podrían haber sido indirectamente atribuibles a la relación entre la edad y las puntuaciones APM-18. Este modelo definió la puntuación APM-18 como la variable de criterio, con las variables predictoras ordenadas en edad y luego sexo. El modelo jerárquico fue diseñado para permitir que la edad absorba tanta varianza como sea posible, con el sexo introducido en el modelo sólo después. Usando este modelo, ambos GLM indicaron un efecto significativo para la edad ( $F = 15,39, p < .001$ ) y luego también para el sexo después de la edad había sido controlado estadísticamente ( $F = 10,66, p = .001$ ).

## Discussion

Los resultados presentados aquí sugieren que el APM-18 puede servir como un compromiso útil entre la menor confiabilidad APM-12 y el APM-36 mucho más largo. Los GLM jerárquicos identifican tanto la edad como el sexo para ser predictores significativos de las puntuaciones APM-18, con individuos y hombres más jóvenes generalmente puntuando más alto. Estos resultados son consistentes con muchos estudios anteriores que examinan la inteligencia general (por ejemplo, Jackson & Rushton, 2006). Los resultados del Estudio 1, sin embargo, no prueban la validez convergente de esta escala en relación con otras medidas de inteligencia. El estudio 2 fue diseñado para hacer esto.

## ESTUDIO 2: VALIDEZ CONVERGENTE

Se llevó a cabo el estudio 2 para evaluar la validez convergente del APM-18 con otras medidas de inteligencia, rendimiento académico y personalidad. Para ello, probamos dos submuestras separadas (n.o 193 y 229) tomadas del estudio 1, cada una de las cuales utilizó diferentes medidas de criterio. En la Muestra 1, se utilizaron dos medidas ampliamente utilizadas de inteligencia para adultos: la Escala de Vocabulario Mill-Hill-Múltiples Conjuntos de Opciones A & B (MHV-MC; J. Raven, Raven, & Court, 1997), desarrollado para ser utilizado en conjunto con el APM-36 como una medida de la capacidad reproductiva, es decir, la capacidad de almacenar y recuperar información (J. C. Raven, 1989); y el Instituto Shipley de Escala Viviente (SILS; Zachary, 1986), que es una prueba de inteligencia independiente compuesta por dos subescalas: Vocabulary, que prueba inteligencia cristalizada, y Abstraction, que prueba la inteligencia

fluida. Además, examinamos el rendimiento académico a través de los puntajes de promedio de calificaciones (GPA) y de prueba de aptitud escolar (SAT, por sus propios informes). En la Muestra 2, examinamos las correlaciones entre las puntuaciones APM-18 y las dimensiones de personalidad de los Cinco Grandes evaluadas con la Escala NEO Five-Factor Inventory (NEO-FFI) (Costa & McCrae, 1992), y la creatividad verbal y de dibujo (Miller & Tal, 2007). Además, las puntuaciones ACT se recogieron en esta segunda muestra.

## **Método**

### ***Participantes***

La muestra 1 estaba compuesta por 193 estudiantes (94 hombres, 99 mujeres) de un curso de psicología introductoria en la Universidad de Arizona. La edad media de los participantes fue de 19,11 años, SD a 1,62 (Mmacho 19,23,SD masculino a 1,07; Mhembra 19,01, SDhembra 2,00). Debido al tiempo requerido para administrar el APM-18, el Shipley y el Mill-Hill, 10 participantes no completaron la Prueba Mill-Hill. Instamos a los participantes a que registraran sus puntuaciones de SAT, ACT y GPA sólo si estaban seguros de ellas; debido a esta restricción, muchas de estas puntuaciones también faltaban.

La muestra 2 estaba compuesta por 229 estudiantes (65 hombres, 164 mujeres) de varios cursos de pregrado en la Universidad de Nuevo México. La edad media de los participantes fue de 20,19 años, SD a 3,43 (Mmale a 21,05,SD masculino a 5,01; Mhembra 19,85, SDfemale á 2,48). Una vez más, instamos a los participantes a registrar sus puntuaciones de ACT sólo si estaban seguros de ellas, dejándonos con puntuaciones de ACT para sólo 129 participantes.

## **Medidas**

### ***APM-18***

El APM-18 consistía en los mismos elementos identificados en el Estudio 1. En la Muestra 1, la forma se presentó primero en una serie de medidas que examinan la inteligencia de adultos. En la Muestra 2, se presentó en medio de un paquete de cuestionarios sobre personalidad, creatividad, comportamiento sexual e inteligencia.

### ***Los SILS***

El SILS (Zachary, 1986) es una medida de autoinuncia de 10 minutos por subescala, de 60 elementos que examina tanto la inteligencia verbal (40 elementos) como la inteligencia abstracta (20 elementos). El examen se considera apropiado para individuos de habla inglesa promedio de 14 a edades adultas, que son tomadores de exámenes motivados. Las validezes y normas publicadas en el manual fueron tomadas de una muestra de 322 reclutas del ejército. Las reliabilidades de la mitad dividida para cada subescala se notifican como .87 para vocabulario, .89 para abstracción y .92 para la puntuación total.

### ***El MHV-MC***

El MHV-MC (J. Raven et al., 1997) es una prueba de vocabulario autoadministrado de 68 elementos diseñada para complementar el APM-36. Mientras que el APM tenía como objetivo medir la capacidad de un individuo para resolver problemas novedosos y pensar de maneras novedosas (es decir, inteligencia fluida), el Mill-Hill tenía como objetivo medir la capacidad de un individuo para recordar información aprendida (es decir, inteligencia cristalizada). En este sentido, indica los logros educativos, los antecedentes culturales y la familiaridad con el lenguaje de la prueba. El Mill-Hill normalmente muestra reliabilidades a la mitad divididas más de .90 y reequipen las reliabilidades de prueba y reprobación entre .87 y .95 (Raven et al., 1997).

## ***Rendimiento académico***

El rendimiento académico se midió por los GPA autoinformados y los puntajes del SAT en la Muestra 1. Se pidió a los participantes de la muestra 2 los puntajes del SAT y del ACT. Una variedad de estudios



han identificado correlaciones moderadas a fuertes entre estas medidas de rendimiento académico y aptitud, y una variedad de otros rasgos, incluyendo inteligencia, personalidad y psicopatología (Barton, Dielman, & Cattell, 1971; Brown, 1994; Dyer, 1987; Mouw & Khanna, 1993).

### **NEO-FFI**

El NEO-FFI (Costa & McCrae, 1992) es la medida más utilizada en la investigación sobre el modelo de personalidad Five-Factor. Es una versión abreviada del Inventario de Personalidad NEO revisado de 240 artículos (NEO-PI-R; Costa & McCrae, 1992), compuesto por 60 elementos que miden cinco factores globales de la personalidad (12 elementos por factor): Apertura a la experiencia, Conciencia, Extraversión, Concenacidad y Neuroticismo. En nuestra versión, los participantes calificaron el grado de acuerdo con declaraciones sobre sus personalidades y propensión conductual en una escala de 5 puntos que va desde el número 2 (fuertemente en desacuerdo) a 0 (neutral) a +2 (fuertemente de acuerdo). Esta escala ha mostrado un IC fuerte, con los alfas de Cronbach que oscilan entre .74 y .89, para cada factor, y una validez intercultural consistente (McCrae & Costa, 1997).

### **Tareas verbales y de creatividad de dibujo**

Los participantes completaron seis tareas de creatividad verbal de 2 minutos y ocho tareas de creatividad de dibujo de 1 minuto (Miller & Tal, 2007). Debido a que un mentalidad orientado al apareamiento promueve la creatividad (Griskevicius, Cialdini, & Kenrick, 2006), se pidió a los participantes que completaran estas tareas de la manera más creativa posible con la intención de atraer a un compañero romántico. Algunos ejemplos de tareas verbales fueron escribir respuestas a preguntas que provocan pensamientos, como "¿Cómo mantendrías un matrimonio emocionante después de los primeros dos años?" "¿Cómo esperas que sea el mundo en 100 años?" y "Imagina que todas las nubes tenían cuerdas muy largas colgando de ellas, cuerdas de cientos de pies de largo. ¿Cuáles serían las implicaciones de ese hecho para la naturaleza y la sociedad?" Había dos tipos de tareas de dibujo, cuatro abstractas (por ejemplo, "Por favor, dibuje un símbolo abstracto, un patrón o una composición que represente su felicidad como niño haciendo una actividad favorita") y cuatro representaciones (por ejemplo, "En el espacio de abajo, dibuje un animal que admire por su fuerza, gracia, velocidad o belleza"). Las respuestas de cada participante a cada una de las 14 tareas de creatividad fueron puntuadas de forma independiente por cuatro calificadores en una escala de creatividad de 1 a 5 puntos. Las medidas de creatividad verbal compuesta y creatividad de dibujo compuestas resultantes mostraron una alta fiabilidad interrater e IC (alfas de Cronbach .91 y .90, respectivamente; Miller & Tal, 2007).

## **Resultados**

### **Muestra 1**

Las estimaciones de IC se calcularon utilizando el alfa de Cronbach. El APM-18 mostró un IC moderado (o .71), con el APM-12 incrustado produciendo un valor ligeramente inferior (o .63). Aunque estas consistencias internas son más bajas que las reportadas en el Estudio 1, siguen siendo moderadas en fuerza.

La puntuación media de APM-18 fue de 10,68, SD a 3,25 (Mmacho a 11,07,SD macho a 3,13; Mhembra 10,31, SDhembra 3,34), con un rango de 13 (cuatro a 17). No hubo ninguna relación entre la puntuación y la edad de APM-18 ( $r = .03$ ,  $p = .71$ ) o el sexo ( $r = .11$ ,  $p = .10$ ). Las puntuaciones medias de cada una de las medidas de inteligencia se pueden ver en el Cuadro 2. Debido a la diferencia de sexo significativa entre las puntuaciones de APM-18 en el Estudio 1, las diferencias de sexo medias en todas las medidas de este estudio se comprobaron mediante pruebas t. No hubo diferencias de sexo significativas para ninguna de las medidas de inteligencia en esta muestra, excepto por una ventaja masculina moderada en las puntuaciones de SAT autoinformadas ( $t = 3,00$ ,  $p = .003$ ). Por lo tanto, los análisis restantes se realizaron en la muestra completa y no por sexo.

Como se ve en el Cuadro 3, tanto el APM-18 como el APM-12 integrado se correlacionaron significativamente con la mayoría de las otras medidas de inteligencia y rendimiento académico y aptitud utilizadas en esta muestra. Específicamente, ambas escalas APM se correlacionan positiva y significativamente con la escala de abstracción de Shipley y los puntajes del SAT de autoinuncia. Esto no

es sorprendente, ya que el APM está diseñado para ser una medida de *g*, que puede identificarse más fácilmente en relación con medidas analíticas abstractas, de las cuales la abstracción Shipley es una, y el SAT contiene una subescala analítica.

### ***Muestra 2***

Las estimaciones de IC se calcularon de nuevo utilizando el alfa de Cronbach. Al igual que en el Estudio 1 y la Muestra 1 de este estudio, el APM-18 mostró una fiabilidad moderada ( $\alpha = .79$ ), mientras que el APM-12 integrado vuelve a mostrar una fiabilidad ligeramente inferior ( $\alpha = .74$ ). La puntuación media de APM-18 fue de 9,53, SD a 3,57 (Mmacho a 10,29, SD macho a 3,98; Mhembra 9,23, SDhembra 3,36), con un rango de 1 a 18. No hubo ninguna relación entre la puntuación de APM-18 y la edad ( $r = .03$ ,  $p = .61$ ), pero hubo entre la puntuación de APM-18 y el sexo ( $r = .13$ ,  $p = .04$ ), con los machos de nuevo anotando ligeramente más alto ( $t = 2.04$ ,  $p < .05$ ). No hubo otras diferencias de sexo en las otras medidas de inteligencia (para todas las *ts* a 1,20,  $p > .05$ ). La Tabla 4 muestra las puntuaciones medias de las medidas de inteligencia y los factores NEO-FFI.

Como se ve en el Cuadro 5, tanto las puntuaciones APM-18 como las APM-12 incrustadas estaban significativamente relacionadas positivamente con la creatividad verbal ( $r_s = .36$  y  $.32$ ,  $p < .001$ , respectivamente), creatividad de dibujo ( $r_s = .34$  y  $.29$ ,  $p < .001$ , respectivamente), y puntuación ACT autoinformada ( $r_s = .44$  y  $.45$ ,  $p < .001$ , respectivamente). Además, las escalas de APM estaban positivamente relacionadas con la apertura ( $r_s = .26$  y  $.25$ ,  $p < .01$ , respectivamente) y negativamente con la conciencia (para cada uno,  $r = .16$ ,  $p < .05$ ), un hallazgo consistente con la investigación anterior (Ackerman & Heestad, 1997; Moutafi, Furnham, & Crump, 2003).

## **DISCUSIÓN**

Cada muestra del estudio 2 utilizó diferentes métodos para evaluar la validez convergente del APM-18. La muestra 1 se centró en las relaciones entre el APM-18 y otras medidas estándar de inteligencia y rendimiento académico (por ejemplo, pruebas de inteligencia verbal, GPA autoinformados y puntuaciones del SAT); mientras que la Muestra 2 examinó la relación entre el APM-18, la creatividad, las puntuaciones de ACT autoinformadas y los rasgos de personalidad de los Cinco Grandes. Ambos estudios confirmaron que el APM-18 está relacionado con estas medidas de manera predecible. En términos generales, tanto el APM-18 como el APM-12 incrustado mostraron el mismo patrón de correlaciones con las otras medidas utilizadas en estos estudios. Sin embargo, el IC más alto del APM-18 sugiere que puede ser mejor en la detección de variaciones individuales en *g*.

## **CONCLUSIÓN**

Cada uno de los 18 elementos utilizados en esta nueva prueba APM-18 fue elegido para mantener la dificultad progresiva tanto de la forma larga (APM-36) como de la forma corta (APM-12). Como era de esperar, aunque la fiabilidad del APM-18 era menor que la del APM-36, era mayor que la del APM-12 desarrollado por Arthur y Day (1994). Además, los patrones de correlación con otras medidas de inteligencia son prácticamente idénticos al APM-12, que, en estudios anteriores, ha demostrado imitar los resultados del APM-36 (Arthur & Day, 1994; Arthur et al., 1999). Combinado con un tiempo medio de administración de 17,53 min (máximo de 25 min), estos hallazgos sugieren que el APM-18 puede funcionar bien como un compromiso para los investigadores que quieren una medida bastante precisa de inteligencia general en un período bastante corto de tiempo. La validación cruzada en las tres muestras notificadas aquí es un intento inicial de recopilar datos normativos para el APM-18. Nuestros resultados pueden generalizar solo a otros estudiantes universitarios. Sin embargo, el corto tiempo de administración del APM-18, su alto IC, validez razonable y facilidad de administración por papel y lápiz en grandes entornos de aulas universitarias lo hacen ideal para estudios de ciencias del comportamiento donde los investigadores quieren una puntuación de inteligencia razonablemente rápida y precisa como parte de una batería de cuestionario más grande.

## **TRANSLATED VERSION: FRENCH**

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

## **VERSION TRADUITE: FRANÇAIS**

Voici une traduction approximative des idées présentées ci-dessus. Cela a été fait pour donner une compréhension générale des idées présentées dans le document. Veuillez excuser toutes les erreurs grammaticales et ne pas tenir les auteurs originaux responsables de ces erreurs.

## **INTRODUCTION**

Le Ravens Progressive Matrices Test, développé par Raven (1941) comme une mesure de l'intelligence générale ( $g$ ), a subi de nombreuses révisions, allant de versions colorées pour les enfants à la matrice standard et avancée pour les adultes de différents niveaux cognitifs. La version la plus récente est la Matrices progressive avancées de Ravens (APM ; Raven, Raven et Court, 1993), qui a été développé pour les populations adultes à plus grande capacité (c.-à-d. au niveau collégial et supérieur). Ce test est construit de 36 éléments de difficulté croissante divisés en trois ensembles de 12 points; dans chaque élément, l'examineur est invité à compléter un modèle visuel en choisissant l'une des huit solutions possibles.

En raison de son format non verbal, l'APM est censé être une mesure culturellement juste et impartiale de l'intelligence fluide (Cattell, 1963), de la capacité éducative (J. Raven et al., 1993), ou, comme nous le mentionnerons, de l'intelligence générale ( $g$ ; Spearman, 1927), et s'est montré particulièrement utile dans les situations où l'anglais n'est pas la langue principale d'un individu. À ce titre, les matrices progressives standard et avancées ont été largement utilisées dans de nombreux contextes appliqués aux États-Unis (p. ex., Ackerman, 1992) et dans de nombreuses cultures (Owen, 1992; J. C. Raven, 2000; Rushton, Cvorovic et Bons, 2007). Cependant, les aspects positifs de ce test sont entachés par son long temps d'administration (40-60 min), ce qui rend difficile l'utilisation dans les milieux multivariés ou de salle de classe limités dans le temps.

En réponse à ces diverses limitations, Arthur et Day (1994) ont développé une forme courte de 12 points de l'APM (que nous appelons APM-12), avec un temps d'administration de 15 min. Plusieurs études ont montré que ce formulaire de 12 points montre des propriétés psychométriques acceptables (p. ex., alpha de Cronbach, fiabilité test-retest, validité convergente; voir Arthur, Tubre, Paul, & Sanchez-Ku, 1999, pour examen). Toutefois, cette forme courte montre une cohérence interne relativement faible et variable (IC). Par exemple, les alphas de Cronbach vont de 0,58 à 0,66 pour la forme courte elle-même et de 0,72 à 0,73 pour les 12 articles de forme courte extraits de la version complète de 36 articles (Arthur & Day, 1994).

Plus récemment, Hamel et Schmittmann (2006) ont fait valoir que l'APM complet de 36 points peut être administré sous forme d'un test de vitesse de 20 minutes. Les scores sur cette forme rapide de l'APM montrent de fortes corrélations avec les scores sur des versions chronométrées plus lentes (40 min,  $r = 0,74$ ) et non chronométrées ( $r = 0,75$ ) de l'APM. Cependant, ces auteurs n'ont pas signalé l'IC de l'échelle de test de vitesse. Nous soupçonnons également que donner aux adultes typiques seulement 20 min pour compléter 36 problèmes de raisonnement abstrait très difficiles pourrait imposer un stress indu.

Le but de la présente étude était d'élaborer une version de forme moyenne de l'APM qui a donné lieu à des IC plus élevés que la version en 12 points (APM-12), mais moins de temps d'administration que l'APM complète de 36 points (APM-36) — une combinaison de fonctionnalités qui pourraient être utiles pour les situations de temps limités et de test de masse. Ici, nous rapportons le développement et la validité de construction de cette échelle de 18 points.

## **ÉTUDE 1 : CONSTRUCTION À L'ÉCHELLE ET VALIDITÉ DE LA CONSTRUCTION**

## **Méthode**

### ***Participants***

Un total de 633 étudiants (198 hommes, 435 femmes) de trois universités du sud-ouest ont participé à cette étude comme une exigence partielle pour le crédit de cours expérimental. L'âge moyen des participants était de 20,92 ans, SD = 4,07 (Mmale = 20,85, SDmale = 3,90; Mfemale = 20,96, SDfemale = 4,15). Les âges variaient de 17 à 58 ans (mâle = 18-41 ans, femelle = 17-58).

### **Mesures**

#### ***The Ravens Advanced Progressive Matrices 18-Item Short Form (APM-18)***

Cette version abrégée de 18 articles de l'APM est imprimée au format livret sur papier blanc de 81/2 po par 11 po, chaque élément d'essai étant imprimé sur une page distincte. Les quatre premières pages du livret de test contiennent trois exemples d'éléments (articles de pratique 1, 5 et 9 de l'APM-36) pour expliquer la tâche.

Les 18 éléments d'essai réels ont été dérivés en ajoutant six éléments de la version plus longue de 36 articles (J. Raven et al., 1993) à Arthur and Day's (1994) version publiée en 12 points. Arthur et Day ont utilisé les articles 1, 4, 8, 11, 15, 18, 21, 23, 25, 30, 31 et 35 de l'APM de 36 points sur la base d'un ensemble de trois règles de décision, qui peuvent être résumées comme a) divisant l'APM en 12 sections en trois points en fonction de la difficulté; b) prendre l'article avec la corrélation point-total la plus élevée pour chaque section; c) dans le cas d'une égalité, y compris l'article qui a entraîné la plus forte baisse d'IC si elle était exclue de l'essai complet. Suivant ces mêmes règles, nous avons ajouté six autres éléments de difficulté croissante : deux autres qui étaient faciles (96 % et 75 % des examinateurs de l'échantillon normatif répondu correctement), deux qui étaient modérés (50 % et 48 % des examinateurs de l'échantillon normatif répondu correctement), et deux qui étaient difficiles (37 % et 32 % des examinateurs de l'échantillon normatif répondu). Ces articles (2, 20, 22, 24, 34 et 32) ont été intégrés de difficulté pour imiter leur ordre de présentation dans l'APM original.

### ***Procédure***

Le nouveau test APM-18 a été donné en classe avec plusieurs examens à la fois. Cela a été fait parce que ce test a été développé comme une mesure de g qui pourrait être utilisé dans des environnements tels que les salles de classe, où il ya des limites de temps sur les séances de recherche. Dans un sous-échantillon (n = 175), des tests ont été effectués sans contraintes de temps, mais avec des délais d'achèvement enregistrés, afin de déterminer le temps moyen nécessaire à l'achèvement. Les deux autres sous-échantillons (n = 232 et n = 226) ont été contraints de terminer l'essai dans un délai de 25 minutes, sans que les temps d'achèvement individuels ne soient enregistrés.

### ***Analyses***

Toutes les analyses statistiques ont été effectuées à l'aide de la version 8.2 de la SAS (Institut SAS, 1999). Les alphas de Cronbach et les corrélations bivariées ont été calculés à l'aide de la procédure PROC CORR. Les tests pour les différences moyennes entre les sexes ont été calculés au moyen de procédures d'essai t (PROC TTEST). Des modèles linéaires généraux hiérarchiques (GLM) ont été testés à l'aide de PROC GLM.

### **Résultats**

Les estimations d'IC ont été calculées en utilisant l'alpha de Cronbach. L'IC de l'échelle APM-18 a donné une fiabilité modérée ( $\alpha = 0,79$ ). Cet alpha est inférieur aux rapports IC normatifs pour l'APM-36 ( $\alpha = 0,84$ ; Forbes, 1964), mais plus élevé que ceux de l'APM-12 (allant de  $\alpha = 0,58-0,66$ ; voir Arthur et al., 1999). En outre, l'alpha de l'APM-18 était plus grand que celui de l'APM-12 incorporé ( $\alpha = 0,73$ ). Le tableau 1 affiche les résultats pour chacun des éléments APM-18, en ce qui concerne leurs corrélations entre l'élément, les difficultés d'élément et l'échelle  $\alpha$  de l'échelle globale si l'élément est supprimé. Comme on

l'a vu, la suppression de tout élément réduit la fiabilité globale de l'échelle, ce qui suggère que tous les éléments doivent être conservés.

Le score moyen apm-18 était de 9,73, SD = 3,59 (M mâle = 10,43, mâle SD = 3,52; M femelle = 9,41, femelle SD = 3,59), avec une gamme de 18. Pour le sous-échantillon dans lequel les temps d'achèvement ont été enregistrés (n = 175), le temps moyen d'achèvement des essais était de 17,5 min (DS = 4,67), avec une plage de 7 à 25 min; 21% des participants ont pris plus de 20 min, mais personne n'a pris plus de 25 min. Dans ce sous-échantillon, il y avait une relation positive significative entre le temps qu'il a fallu aux participants pour passer le test et leur score APM-18 ( $r = 0,41$ ,  $p < .001$ ), mais il n'y avait pas de relation entre l'âge et la note APM-18, ou l'âge et le temps requis pour terminer le test (pour chaque  $r \leq .03$ ,  $p \geq .71$ ). Toutefois, pour l'échantillon complet (N = 633), les jeunes participants ont obtenu un score un peu plus élevé (âge et scores APM-18 corrélés  $r = -.15$ ,  $p < .001$ ), et les hommes ont obtenu une note un peu plus élevée — sexe (femelle = 0, mâle = 1) et APM-18 scores corrélés  $r = 0,13$ ,  $p < .001$ .

Des GLM hiérarchiques ont été testés pour déterminer si les différences apparentes entre les scores APM-18 masculins et féminins auraient pu être indirectement attribuables à la relation entre l'âge et les scores APM-18. Ce modèle a défini le score APM-18 comme la variable de critère, les variables prédictorales ordonnées étant l'âge, puis le sexe. Le modèle hiérarchique a été conçu pour permettre à l'âge d'absorber autant de variance que possible, avec le sexe entré dans le modèle seulement après. À l'aide de ce modèle, les deux GLM ont indiqué un effet significatif pour l'âge ( $F = 15,39$ ,  $p < 0,001$ ) et ensuite aussi pour le sexe après l'âge avait été contrôlé statistiquement ( $F = 10,66$ ,  $p = 0,001$ ).

## Discussion

Les résultats présentés ici suggèrent que l'APM-18 peut servir de compromis utile entre la fiabilité inférieure APM-12 et l'APM-36 beaucoup plus long. Les GLM hiérarchiques identifient à la fois l'âge et le sexe comme des prédicteurs significatifs des scores apm-18, les individus plus jeunes et les hommes obtenant généralement des notes plus élevées. Ces résultats concordent avec de nombreuses études antérieures portant sur l'intelligence générale (p. ex., Jackson et Rushton, 2006). Les résultats de l'étude 1 ne testent toutefois pas la validité convergente de cette échelle par rapport à d'autres mesures de l'intelligence. L'étude 2 a été conçue pour ce faire.

## ÉTUDE 2 : VALIDITÉ CONVERGENTE

L'étude 2 a été menée pour évaluer la validité convergente de l'APM-18 avec d'autres mesures de l'intelligence, de la réussite scolaire et de la personnalité. Pour ce faire, nous avons testé deux sous-échantillons distincts (n = 193 et 229) tirés de l'étude 1, dont chacun utilisait des mesures de critères différentes. Dans l'échantillon 1, deux mesures largement utilisées de l'intelligence adulte ont été utilisées : l'échelle de vocabulaire mill-hill-ensembles à choix multiples A & B (MHV-MC; J. Raven, Raven et Court, 1997), qui a été conçu pour être utilisé conjointement avec l'APM-36 comme mesure de la capacité de reproduction, c'est-à-dire la capacité de stocker et de récupérer de l'information (J. C. Raven, 1989); et le Shipley Institute of Living Scale (SILS; Zachary, 1986), qui est un test d'intelligence autonome composé de deux sous-échelles : le vocabulaire, qui teste l'intelligence cristallisée, et l'abstraction, qui teste l'intelligence fluide. De plus, nous avons examiné le rendement scolaire par l'entremise de la moyenne des points de grade (GPA) et des résultats des tests d'aptitude scolaire (SAT). Dans l'échantillon 2, nous avons examiné les corrélations entre les scores APM-18 et les dimensions de la personnalité des cinq grands évalués à l'échelle de l'inventaire à cinq facteurs NEO (NEO-FFI) (Costa et McCrae, 1992) et la créativité verbale et dessin (Miller et Tal, 2007). De plus, les scores de l'ACT ont été recueillis dans ce deuxième échantillon.

## **Méthode**

### ***Participants***

L'échantillon 1 était composé de 193 étudiants (94 hommes, 99 femmes) d'un cours d'introduction en psychologie à l'Université de l'Arizona. L'âge moyen des participants était de 19,11 ans, SD = 1,62 (M mâle = 19,23, SD mâle = 1,07; Mfemelle = 19,01, femelle SD = 2,00). En raison du temps nécessaire à l'administration de l'APM-18, du Shipley et du Mill-Hill, 10 participants n'ont pas terminé le test Mill-Hill. Nous avons exhorté les participants à enregistrer leurs notes sat, ACT et GPA seulement s'ils étaient certains d'entre eux; en raison de cette contrainte, beaucoup de ces scores manquaient également.

L'échantillon 2 était composé de 229 étudiants (65 hommes, 164 femmes) de divers cours de premier cycle à l'Université du Nouveau-Mexique. L'âge moyen des participants était de 20,19, SD = 3,43 (Mmâle = 21,05, SD mâle = 5,01; Mfemelle = 19,85, SDfemale = 2,48). Encore une fois, nous avons exhorté les participants à enregistrer leurs scores ACT seulement s'ils étaient certains d'entre eux, nous laissant avec des scores ACT pour seulement 129 participants.

## **Mesures**

### ***APM-18***

L'APM-18 se composait des mêmes éléments identifiés dans l'étude 1. Dans l'échantillon 1, le formulaire a été présenté en premier dans une série de mesures examinant l'intelligence des adultes. Dans l'échantillon 2, il a été présenté au milieu d'un paquet de questionnaires concernant la personnalité, la créativité, le comportement sexuel et l'intelligence.

### ***Le SILS***

Le SILS (Zachary, 1986) est une mesure chronométrée (10 min par sous-échelle), une mesure d'auto-déclaration de 60 points qui examine à la fois l'intelligence verbale (40 éléments) et l'intelligence abstraite (20 points). Le test est considéré comme approprié pour les personnes anglophones moyennes de 14 à l'âge adulte, qui sont des personnes motivées. Les validités et les normes publiées dans le manuel ont été tirées d'un échantillon de 322 recrues de l'armée. Les reliabilités fractionnées pour chaque sous-échelle sont signalées comme 0,87 pour Vocabulary, .89 pour Abstraction et .92 pour le score total.

### ***Le MHV-MC***

Le MHV-MC (J. Raven et al., 1997) est un test de vocabulaire à choix multiples auto-administré de 68 éléments conçu pour compléter l'APM-36. Alors que l'APM visait à mesurer la capacité d'une personne à résoudre de nouveaux problèmes et à penser de nouvelles façons (c.-à-d. l'intelligence fluide), le Mill-Hill visait à mesurer la capacité d'une personne à se souvenir de l'information apprise (c.-à-d. l'intelligence cristallisée). Dans cette mesure, il indique le niveau d'instruction, les antécédents culturels et la familiarité avec la langue du test. Le Mill-Hill montre généralement des reliabilités à moitié fractionnée de plus de 0,90 et des reliabilités test-retest variant entre 0,87 et 0,95 (Raven et al., 1997).

### ***Performance académique***

Le rendement scolaire a été mesuré par les GPA auto-déclarés et les scores sat dans l'échantillon 1. On a demandé aux participants de l'échantillon 2 des notes sat et act. Diverses études ont identifié des corrélations modérées à fortes entre ces résultats scolaires et les mesures d'aptitude, et une variété d'autres traits, y compris l'intelligence, la personnalité et la psychopathologie (Barton, Dielman et Cattell, 1971; Brown, 1994; Dyer, 1987; Mouw et Khanna, 1993).

### ***NEO-FFI***

Le NEO-FFI (Costa & McCrae, 1992) est la mesure la plus largement utilisée dans la recherche sur le modèle de personnalité à cinq facteurs. Il s'agit d'une version raccourcie de l'inventaire révisé de personnalité NEO de 240

articles (NEO-PI-R; Costa & McCrae, 1992), composé de 60 éléments qui mesurent cinq facteurs de personnalité mondiale (12 éléments par facteur) : Ouverture à l'expérience, conscience, extraversion, amabilité et névrosisme. Dans notre version, les participants ont noté le degré d'accord avec les déclarations sur leur personnalité et leurs propensions comportementales sur une échelle de 5 points allant de -2 (fortement en désaccord) à 0 (neutre) à +2 (tout à fait d'accord). Cette échelle a montré ic forte, avec les alphas de Cronbach allant entre .74 à .89, pour chaque facteur, et la validité interculturelle cohérente (McCrae & Costa, 1997).

### **Tâches de créativité verbales et de dessin**

Les participants ont accompli six tâches de créativité verbale de 2 minutes et huit tâches de créativité de dessin d'une minute (Miller et Tal, 2007). Parce qu'un état d'esprit axé sur l'accouplement favorise la créativité (Griskevicius, Cialdini et Kenrick, 2006), les participants ont été invités à accomplir ces tâches de la manière la plus créative possible avec l'intention d'attirer un partenaire romantique. Parmi les exemples de tâches verbales, mentionnons l'écriture de réponses à des questions qui suscitent la réflexion, comme « Comment garderiez-vous un mariage excitant après les deux premières années? » « Qu'espérez-vous que le monde sera dans 100 ans ? » et « Imaginez que tous les nuages avaient de très longues cordes qui leur pendaient — des cordes de plusieurs centaines de pieds de long. Quelles seraient les implications de ce fait pour la nature et la société? Il y avait deux types de tâches de dessin, quatre résumées (p. ex., « S'il vous plaît, dessinez un symbole abstrait, un motif ou une composition qui représente votre bonheur d'enfant faisant une activité préférée ») et quatre représentations (p. ex., « Dans l'espace ci-dessous, veuillez dessiner un animal que vous admirez pour sa force, sa grâce, sa vitesse ou sa beauté »). Les réponses de chaque participant à chacune des 14 tâches de créativité ont été notées indépendamment par quatre évaluateurs sur une échelle de créativité de 1 à 5 points. La créativité verbale composite qui en a résulté et les mesures de créativité de dessin ont montré une grande fiabilité interrater et IC (alphas de Cronbach = 0,91 et 0,90, respectivement; Miller et Tal, 2007).

## **Résultats**

### ***Échantillon 1***

Les estimations d'IC ont été calculées à l'aide de l'alpha de Cronbach. L'APM-18 a montré ic modéré ( $\alpha = .71$ ), avec l'APM-12 incorporé donnant une valeur légèrement inférieure ( $\alpha = .63$ ). Bien que ces consistances internes soient inférieures à celles rapportées dans l'étude 1, elles sont encore modérées en force.

Le score moyen apm-18 était de 10,68, SD = 3,25 (M mâle = 11,07, mâle SD = 3,13; M femelle = 10,31, femelle SD = 3,34), avec une gamme de 13 (quatre à 17). Il n'y avait aucune relation entre le score APM-18 et l'âge ( $r = 0,03$ ,  $p = 0,71$ ), ou le sexe ( $r = 0,11$ ,  $p = 0,10$ ). Les scores moyens pour chacune des mesures de l'intelligence peuvent être vus dans le tableau 2. En raison de la différence significative entre les sexes entre les scores APM-18 dans l'étude 1, les différences moyennes entre les sexes sur toutes les mesures de cette étude ont été vérifiées par des tests t. Il n'y avait aucune différence significative entre les sexes pour l'une ou l'autre des mesures d'intelligence de cet échantillon, à l'exception d'un avantage modéré des hommes sur les scores de SAT autodéclarés ( $t = -3,00$ ,  $p = 0,003$ ). Par conséquent, les analyses restantes ont été effectuées sur l'échantillon complet plutôt que par sexe.

Comme on peut le voir au tableau 3, l'APM-18 et l'APM-12 intégré se sont corrélés de façon significative avec la plupart des autres mesures de l'intelligence et de la réussite scolaire et des aptitudes

utilisées dans cet échantillon. Plus précisément, les deux échelles apm corrélées positivement et significativement plus fortement avec l'échelle Shipley Abstraction et l'auto-rapport scores SAT. Cela n'est pas surprenant car l'APM est conçu pour être une mesure de g, qui peut être plus facilement identifié par rapport à des mesures abstraites et analytiques, dont l'abstraction Shipley est l'un, et le SAT contient une sous-échelle analytique.

### **Exemple 2**

Les estimations d'IC ont de nouveau été calculées à l'aide de l'alpha de Cronbach. Comme dans l'étude 1 et l'échantillon 1 de cette étude, l'APM-18 a montré une fiabilité modérée ( $\alpha = 0,79$ ), tandis que l'APM-12 intégré montre encore une fois une fiabilité légèrement inférieure ( $\alpha = 0,74$ ). Le score moyen apm-18 était de 9,53, SD = 3,57 (M mâle = 10,29, mâle SD = 3,98; M femelle = 9,23, femelle SD = 3,36), avec une fourchette de 1 à 18. Il n'y avait pas de relation entre le score APM-18 et l'âge ( $r = -0,03$ ,  $p = 0,61$ ), mais il y avait entre le score APM-18 et le sexe ( $r = 0,13$ ,  $p = 0,04$ ), les hommes obtenant à nouveau un score légèrement plus élevé ( $t = -2,04$ ,  $p < .05$ ). Il n'y avait pas d'autres différences entre les sexes sur les autres mesures d'intelligence (pour tous les  $t_s \geq -1,20$ ,  $p > .05$ ). Le tableau 4 montre les scores moyens pour les mesures de renseignement et les facteurs NEO-FFI.

Comme on peut le voir au tableau 5, les scores APM-18 et APM-12 intégrés étaient sensiblement liés positivement à la créativité verbale ( $r_s = 0,36$  et  $0,32$ ,  $p < .001$ , respectivement), la créativité de dessin ( $r_s = .34$  et  $.29$ ,  $p < .001$ , respectivement), et le score act auto-déclaré ( $r_s = 0,44$  et  $.45$ ,  $p < .001$ , respectivement). En outre, les échelles apm étaient positivement liées à l'ouverture ( $r_s = 0,26$  et  $0,25$ ,  $p < 0,01$ , respectivement) et négativement à la conscience (pour chacune,  $r = -.16$ ,  $p < .05$ ), une constatation compatible avec les recherches précédentes (Ackerman & Heggestad, 1997; Moutafi, Furnham et Crump, 2003).

## **DISCUSSION**

Chaque échantillon de l'étude 2 a utilisé différentes méthodes pour évaluer la validité convergente de l'APM-18. L'échantillon 1 portait sur les relations entre l'APM-18 et d'autres mesures standard de l'intelligence et de la réussite scolaire (p. ex., tests d'intelligence verbale, DPap autodéclarés et scores sat); tandis que l'échantillon 2 a examiné la relation entre l'APM-18, la créativité, les scores d'ACT autodéclarants et les traits de personnalité des Cinq Grands. Les deux études ont confirmé que l'APM-18 est liée à ces mesures d'une manière prévisible. D'une manière générale, l'APM-18 et l'APM-12 intégré ont montré le même schéma de corrélations avec les autres mesures utilisées dans ces études. Toutefois, l'IC plus élevé de l'APM-18 suggère qu'il pourrait être meilleur pour détecter les variations individuelles dans.

## **CONCLUSION**

Chacun des 18 articles utilisés dans ce nouveau test APM-18 a été choisi pour maintenir la difficulté progressive de la forme longue (APM-36) et de la forme courte (APM-12). Sans surprise, bien que la fiabilité de l'APM-18 ait été inférieure à celle de l'APM-36, elle était supérieure à celle de l'APM-12 développée par Arthur and Day (1994). En outre, les modèles de corrélation avec d'autres mesures de l'intelligence sont pratiquement identiques à l'APM-12, qui a, dans des études précédentes, été montré pour imiter les résultats APM-36 (Arthur & Day, 1994; Arthur et coll., 1999). Combiné avec un temps



d'administration moyen de 17,53 min (25 min maximum), ces résultats suggèrent que l'APM-18 peut fonctionner bien comme un compromis pour les chercheurs qui veulent une mesure assez précise de l'intelligence générale dans un laps de temps assez court. La validation croisée des trois échantillons signalés ici est une première tentative de collecte de données normatives pour l'APM-18. Nos résultats peuvent généraliser uniquement pour les autres étudiants. Cependant, le court temps d'administration de l'APM-18, une forte IC, une validité raisonnable et la facilité d'administration par papier et crayon dans les grandes salles de classe des collèges le rendent idéal pour les études de sciences comportementales où les chercheurs veulent un score d'intelligence raisonnablement rapide et précis dans le cadre d'une batterie de questionnaire plus grande.

### **TRANSLATED VERSION: GERMAN**

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

### **ÜBERSETZTE VERSION: DEUTSCH**

Hier ist eine ungefähre Übersetzung der oben vorgestellten Ideen. Dies wurde getan, um ein allgemeines Verständnis der in dem Dokument vorgestellten Ideen zu vermitteln. Bitte entschuldigen Sie alle grammatikalischen Fehler und machen Sie die ursprünglichen Autoren nicht für diese Fehler verantwortlich.

### **EINLEITUNG**

Der Ravens Progressive Matrices Test, entwickelt von Raven (1941) als Maß für die allgemeine Intelligenz (g), hat viele Überarbeitungen durchlaufen, von farbigen Versionen für Kinder bis hin zu Standard- und fortgeschrittenen Matrizen für Erwachsene unterschiedlicher kognitiver Niveaus. Die neueste Version ist die Ravens Advanced Progressive Matrices (APM; Raven, Raven, & Court, 1993), das für erwachsene Populationen mit höherer Befähigung (d. h. College-Level und höher) entwickelt wurde. Dieser Test besteht aus 36 Elementen mit zunehmendem Schwierigkeitsgrad, die in drei 12-Punkte-Sets unterteilt sind; In jedem Element wird der Prüfer aufgefordert, ein visuelles Muster zu vervollständigen, indem er eine von acht möglichen Lösungen auswählt.

Aufgrund seines nonverbalen Formats soll das APM ein kulturell faires, unvoreingenommenes Maß an fluider Intelligenz (Cattell, 1963), erzieherische Fähigkeit (J. Raven et al., 1993) oder, wie wir es nennen werden, allgemeine Intelligenz (g; Spearman, 1927) und hat sich in Situationen, in denen Englisch nicht die Primäre sprache eines Individuums ist, als besonders nützlich erwiesen. Als solche wurden die Standard und Advanced Progressive Matrices in vielen angewandten Einstellungen in den Vereinigten Staaten (z.B. Ackerman, 1992) und in vielen Kulturen (Owen, 1992; J. C. Raven, 2000; Rushton, Cvorovic, & Bons, 2007). Die positiven Aspekte dieses Tests werden jedoch durch seine lange Verabreichungszeit (40-60 min) beeinträchtigt, was die Verwendung in zeitbeschränkten multivariaten Forschungs- oder Klassenzimmerumgebungen erschwert.

Als Antwort auf diese verschiedenen Einschränkungen entwickelte Arthur and Day (1994) eine 12-Punkte-Kurzform des APM (die wir APM-12 nennen), mit einer Verabreichungszeit von 15 min. Mehrere Studien haben gezeigt, dass dieses 12-Punkte-Formular akzeptable psychometrische Eigenschaften aufweist (z.B. Cronbachs Alpha, Test-Retest-Zuverlässigkeit, konvergente Gültigkeit; siehe Arthur, Tubre, Paul, & Sanchez-Ku, 1999, zur Überprüfung). Diese Kurzform weist jedoch eine relativ geringe und variable interne Konsistenz (IC) auf. Cronbachs Alphas reichen beispielsweise von .58 bis .66 für die

Kurzform selbst und .72 bis .73 für die 12 Kurzformelemente, die aus der vollständigen 36-Elemente-Version extrahiert wurden (Arthur & Day, 1994).

In jüngerer Zeit haben Hamel und Schmittmann (2006) argumentiert, dass der komplette 36-Punkte-APM als 20-min-Geschwindigkeitstest verabreicht werden kann. Die Ergebnisse dieser beschleunigten Form des APM zeigen starke Korrelationen mit Partituren auf langsameren Zeit (40 min,  $r = 0,74$ ) und unzeitierten Versionen ( $r = 0,75$ ) des APM. Diese Autoren konnten jedoch den IC der Geschwindigkeitstestskaala nicht melden. Wir vermuten auch, dass es zu unzumutbarem Stress führen könnte, wenn typische Erwachsene nur 20 min sind, um 36 sehr herausfordernde abstrakte Argumentationsprobleme zu erledigen.

Der Zweck der aktuellen Studie war die Entwicklung einer Mittelformversion des APM, die zu einem höheren IC als der 12-Punkte-Version (APM-12) führte, aber kürzere Administrationszeit als das vollständige 36-Punkte-APM (APM-36) – eine Kombination von Features, die für zeitbeschränkte und Massentestsituationen nützlich sein könnten. Hier berichten wir über die Entwicklung und Die Konstruktgültigkeit dieser 18-Punkte-Skala.

## **STUDIE 1: MAßSTABSBAU UND BAUGÜLTIGKEIT**

### **Methode**

#### ***Teilnehmer***

Insgesamt nahmen 633 Studierende (198 männlich, 435 weiblich) von drei südwestlichen Universitäten an dieser Studie als Teilanforderung für experimentelle Studienkredite teil. Das Durchschnittsalter der Teilnehmer betrug 20,92, SD = 4,07 (Mmännlich = 20,85, SDmale = 3,90; Mweiblich = 20,96, SDweiblich = 4,15). Das Alter reichte von 17 bis 58 Jahren (männlich = 18-41, weiblich = 17-58 Jahre).

#### **Maßnahmen**

##### ***Die Ravens Advanced Progressive Matrices 18-Item Kurzform (APM-18)***

Diese 18-teilige Kurzformversion des APM wird in einem Booklet-Format auf 81"-Weißpapier gedruckt, wobei jeder Testartikel auf einer separaten Seite gedruckt wird. Die ersten vier Seiten der Testbroschüre enthalten drei Beispielelemente (Übungspunkte 1, 5 und 9 aus APM-36), um die Aufgabe zu erläutern.

Die 18 tatsächlichen Testpunkte wurden abgeleitet, indem sechs Elemente aus der längeren 36-Elemente-Version (J. Raven et al., 1993) zu Arthur and Day es (1994) veröffentlichten 12-Elemente-Version hinzugefügt wurden. Arthur und Day verwendeten die Artikel 1, 4, 8, 11, 15, 18, 21, 23, 25, 30, 31 und 35 aus dem 36-Punkte-APM basierend auf einer Reihe von drei Entscheidungsregeln, die als a) Aufteilung der APM in 12, dreiteilige Abschnitte basierend auf Schwierigkeiten zusammengefasst werden können; b) Die Aufnahme des Postens mit der höchsten Punkt-Gesamtkorrelation für jeden Abschnitt; und c) im Falle eines Unentschiedens, einschließlich des Postens, der zu dem größten Rückgang des IC geführt hat, wenn er von der vollständigen Prüfung ausgeschlossen wurde. Nach diesen Regeln haben wir sechs weitere Punkte mit zunehmender Schwierigkeit hinzugefügt – zwei, die einfach waren (96% und 75% der Untersuchten aus der normativen Stichprobe richtig beantwortet), zwei, die moderat waren (50% und 48% der Untersuchten aus der normativen Stichprobe richtig beantwortet) und zwei, die schwierig waren (37% und 32% der Untersuchten aus der normativen Stichprobe richtig beantwortet). Diese Elemente (2, 20, 22, 24, 34 und 32) wurden mit Schwierigkeiten integriert, um ihre Darstellungsreihenfolge im ursprünglichen APM nachzuahmen.

#### ***Verfahren***

Der neue APM-18-Test wurde in Klassenzimmern mit mehreren Tests gleichzeitig gegeben. Dies geschah, weil dieser Test als Maß für g entwickelt wurde, das in Umgebungen wie Klassenzimmern verwendet werden kann, in denen es Fristen für Forschungssitzungen gibt. In einer Unterstichprobe ( $n = 175$ ) wurden Tests ohne Zeitbeschränkungen, aber mit aufgezeichneten Fertigstellungszeiten durchgeführt, um die durchschnittliche Zeit zu bestimmen, die für den Abschluss benötigt wird. Die beiden anderen

Teilstichproben ( $n = 232$  und  $n = 226$ ) wurden gezwungen, den Test innerhalb von 25 min abzuschließen, ohne dass einzelne Fertigstellungszeiten aufgezeichnet wurden.

### **Analysen**

Alle statistischen Analysen wurden mit SAS Version 8.2 (SAS Institute, 1999) durchgeführt. Cronbachs Alphas und bivariate Korrelationen wurden mit dem PROC CORR-Verfahren berechnet. Tests auf mittlere Unterschiede zwischen den Geschlechtern wurden durch t-Testverfahren (PROC TTEST) berechnet. Hierarchische allgemeine lineare Modelle (GLMs) wurden mit PROC GLM getestet.

### **Ergebnisse**

IC-Schätzungen wurden mit Hilfe von Cronbachs Alpha berechnet. Der IC der APM-18-Skala ergab eine moderate Zuverlässigkeit ( $b = .79$ ). Dieses Alpha ist niedriger als normative IC-Berichte für den APM-36 ( $n = .84$ ; Forbes, 1964), aber höher als die für die APM-12 (von  $. = .58-.66$ ; siehe Arthur et al., 1999). Darüber hinaus war das Alpha des APM-18 größer als das des eingebetteten APM-12 ( $b = .73$ ). Tabelle 1 zeigt die Ergebnisse für jedes aPM-18-Element in Bezug auf ihre Positions-Gesamtkorrelationen, Artikelschwierigkeiten und Maßstab skalieren des Gesamtmaßstabs, wenn der Artikel gelöscht wird. Wie bereits zu sehen, verringert das Löschen eines Elements die allgemeine Zuverlässigkeit der Skala, was darauf hindeutet, dass alle Elemente beibehalten werden sollten.

Der mittlere APM-18-Wert betrug 9,73,  $SD = 3,59$  (Männlich = 10,43,  $SD$ -Stecker = 3,52; Weiblich = 9,41,  $SD$ weiblich = 3,59), mit einem Bereich von 18. Für die Teilstichprobe, in der Fertigstellungszeiten aufgezeichnet wurden ( $n = 175$ ), betrug die durchschnittliche Testabschlusszeit 17,5 min ( $SD = 4,67$ ), mit einem Bereich von 7 bis 25 min; 21% der Teilnehmer dauerten länger als 20 min, aber niemand dauerte länger als 25 min. In dieser Unterstichprobe gab es eine signifikante positive Beziehung zwischen der Zeit, die die Teilnehmer für den Test benötigten, und ihrem APM-18-Score ( $r = .41$ ,  $p < .001$ ), aber es gab keine Beziehung zwischen Alter und APM-18-Score oder Alter und Zeit, die für den Abschluss des Tests erforderlich waren (für jeden  $r = 0,03$ ,  $p = .71$ ). In der gesamten Stichprobe ( $N = 633$ ) schnitten die jüngeren Teilnehmer jedoch etwas höher ab (Alter und APM-18-Scores korrelierten  $r = 0,15$ ,  $p < 0,001$ ), und die Männer schnitten etwas höher ab – Geschlecht (weiblich = 0, männlich = 1) und APM-18-Scores korrelierten  $r = 0,13$ ,  $p < 001$ .

Hierarchische GLMs wurden getestet, um zu untersuchen, ob die offensichtlichen Unterschiede in den männlichen und weiblichen APM-18-Scores indirekt auf die Beziehung zwischen Alter und APM-18-Scores zurückzuführen sein könnten. Dieses Modell definierte den APM-18-Score als Kriteriumsvariable, wobei die geordneten Prädiktorvariablen Alter und dann Geschlecht waren. Das hierarchische Modell wurde entwickelt, um das Alter so viel Varianz wie möglich zu absorbieren, wobei Sex erst danach in das Modell aufgenommen wurde. Bei Verwendung dieses Modells zeigten beide GLMs einen signifikanten Effekt für das Alter ( $F = 15,39$ ,  $p < 0,001$ ) und dann auch für das Geschlecht nach dem Alter statistisch gesteuert ( $F = 10,66$ ,  $p = 0,001$ ).

### **Diskussion**

Die hier vorgestellten Ergebnisse deuten darauf hin, dass die APM-18 als nützlicher Kompromiss zwischen der niedrigeren Zuverlässigkeit APM-12 und der viel längeren APM-36 dienen kann. Die hierarchischen GLMs identifizieren sowohl Alter als auch Geschlecht als signifikante Prädiktoren für APM-18-Scores, wobei jüngere Personen und Männer in der Regel höher punkten. Diese Ergebnisse stimmen mit vielen früheren Studien über die allgemeine Intelligenz überein (z. B. Jackson & Rushton, 2006). Die Ergebnisse der Studie 1 testen jedoch nicht die konvergente Gültigkeit dieser Skala im Vergleich zu anderen Messgrößen der Intelligenz. Studie 2 wurde dazu entwickelt.

## **STUDIE 2: KONVERGENTE GÜLTIGKEIT**

Studie 2 wurde durchgeführt, um die konvergente Gültigkeit des APM-18 mit anderen Messgrößen der Intelligenz, akademischen Leistung und Persönlichkeit zu bewerten. Dazu haben wir zwei getrennte Unterstichproben ( $n = 193$  und  $229$ ) aus der Studie 1 getestet, von denen jede unterschiedliche Kriterienmaße verwendete. In Probe 1 wurden zwei weit verbreitete Messgrößen der Erwachsenenintelligenz verwendet: die Mill-Hill Vocabulary Scale–Multiple Choice Sets A & B (MHV-MC; J. Raven, Raven, & Court, 1997), entwickelt, um in Verbindung mit dem APM-36 als Maß für die Fortpflanzungsfähigkeit verwendet zu werden, d. h. die Fähigkeit, Informationen zu speichern und abzurufen (J. C. Raven, 1989); und das Shipley Institute of Living Scale (SILS; Zachary, 1986), ein eigenständiger Intelligenztest, der aus zwei Subskalen besteht– Vocabulary, das kristallisierte Intelligenz testet, und Abstraktion, die Fluidintelligenz testet. Außerdem haben wir die akademische Leistung anhand von selbst gemeldeten Notenpunkten (GPA) und SAT-Scores (Scholastic Aptitude Test) untersucht. In Sample 2 untersuchten wir Korrelationen zwischen APM-18-Scores und Big Five Persönlichkeitsdimensionen, die mit der NEO Five-Factor Inventory (NEO-FFI) Scale (Costa & McCrae, 1992) bewertet wurden, und verbale und zeichnende Kreativität (Miller & Tal, 2007). Darüber hinaus wurden in dieser zweiten Stichprobe ACT-Scores gesammelt.

## **Methode**

### ***Teilnehmer***

Probe 1 bestand aus 193 Studenten (94 männlich, 99 weiblich) aus einem Einführungskurs für Psychologie an der University of Arizona. Das Durchschnittsalter der Teilnehmer betrug 19,11,  $SD = 1,62$  (Mmännlich = 19,23,  $SD$ männlich = 1,07; Mweiblich = 19,01,  $SD$ weiblich = 2,00). Aufgrund der Zeit, die für die Verwaltung der APM-18, des Shipley und des Mill-Hill erforderlich war, absolvierten 10 Teilnehmer den Mill-Hill-Test nicht. Wir forderten die Teilnehmer auf, ihre SAT-, ACT- und GPA-Werte nur dann aufzuzeichnen, wenn sie sich dessen sicher waren; Aufgrund dieser Einschränkung fehlten auch viele dieser Punkte.

Stichprobe 2 bestand aus 229 Studenten (65 männlich, 164 weiblich) aus verschiedenen Bachelor-Studiengängen an der University of New Mexico. Das Durchschnittsalter der Teilnehmer betrug 20,19,  $SD = 3,43$  (Mmale = 21,05,  $SD$ männlich = 5,01; Mweiblich = 19,85,  $SD$ female = 2,48). Auch hier forderten wir die Teilnehmer auf, ihre ACT-Werte nur dann aufzuzeichnen, wenn sie sich sicher waren, so dass wir nur 129 Teilnehmer mit ACT-Scores auskommen.

## **Maßnahmen**

### ***APM-18***

Die APM-18 bestand aus den gleichen Punkten, die in Studie 1 identifiziert wurden. In Probe 1 wurde das Formular zuerst in einer Reihe von Maßnahmen zur Untersuchung der Intelligenz von Erwachsenen vorgestellt. In Probe 2 wurde es in der Mitte eines Fragebogenpakets über Persönlichkeit, Kreativität, Sexualverhalten und Intelligenz vorgestellt.

### ***Die SILS***

Die SILS (Zachary, 1986) ist eine zeitgezeitete (10 min pro Subskala), 60-Punkte-Selbstanzeige-Maßnahme, die sowohl verbale Intelligenz (40 Elemente) als auch abstrakte Intelligenz (20 Elemente) untersucht. Der Test gilt als geeignet für durchschnittlich englischsprachige Personen im Alter von 14 bis zum Erwachsenenalter, die motivierte Testteilnehmer sind. Die im Handbuch veröffentlichten Gültigkeiten und Normen stammen aus einer Stichprobe von 322 Rekruten der Armee. Split-Halb-Reliabilities für jede Teilskala werden als .87 für Vokabeln, 0,89 für Abstraktion und 0,92 für die Gesamtpunktzahl gemeldet.

### ***Der MHV-MC***

Der MHV-MC (J. Raven et al., 1997) ist ein selbstverwalteter Multiple-Choice-Vokabulartest mit 68-Item, der die APM-36 ergänzt. Während das APM darauf abzielte, die Fähigkeit eines Individuums zu messen, neue Probleme zu lösen und auf neuartige Weise zu denken (d. h. flüssige Intelligenz), zielte der Mill-Hill darauf ab, die Fähigkeit eines Individuums zu messen, sich an erlernte Informationen (d. h.

kristallisierte Intelligenz) zu erinnern. Insofern zeigt es Bildungsabschlüsse, kulturellen Hintergrund und Vertrautheit mit der Sprache des Tests an. Der Mill-Hill zeigt in der Regel Split-Halb-Reliabilitäten von mehr als 0,90 und Test-Retest-Reliabilitäten zwischen 0,87 und 0,95 (Raven et al., 1997).

### ***Akademische Leistung***

Die akademische Leistung wurde anhand selbst gemeldeter GPAs und SAT-Scores in Probe 1 gemessen. Die Teilnehmer von Sample 2 wurden nach SAT- und ACT-Scores gefragt. Eine Vielzahl von Studien hat moderate bis starke Korrelationen zwischen diesen akademischen Leistungen und Eignungsmaßnahmen und einer Vielzahl anderer Merkmale identifiziert, einschließlich Intelligenz, Persönlichkeit und Psychopathologie (Barton, Dielman, & Cattell, 1971; Brown, 1994; Dyer, 1987; Mouw & Khanna, 1993).

### ***NEO-FFI***

Der NEO-FFI (Costa & McCrae, 1992) ist die am weitesten verbreitete Maßnahme in der Erforschung des Fünf-Faktor-Modells der Persönlichkeit. Es ist eine gekürzte Version des 240-teiligen überarbeiteten NEO Persönlichkeitsinventars (NEO-PI-R; Costa & McCrae, 1992), bestehend aus 60 Artikeln, die fünf globale Persönlichkeitsfaktoren messen (12 Elemente pro Faktor): Offenheit für Erfahrung, Gewissenhaftigkeit, Extraversion, Angenehmheit und Neurotizismus. In unserer Version bewerteten die Teilnehmer den Grad der Übereinstimmung mit Aussagen über ihre Persönlichkeit und Verhaltensneigungen auf einer 5-Punkte-Skala von 2 (stark widersprechen) bis 0 (neutral) bis +2 (stark einverstanden). Diese Skala hat einen starken IC gezeigt, mit Cronbachs Alphas zwischen 0,74 und 0,89, für jeden Faktor, und konsistente interkulturelle Gültigkeit (McCrae & Costa, 1997).

### ***Verbale und zeichnende Kreativitätsaufgaben***

Die Teilnehmer absolvierten sechs zwei-minütigen verbale Kreativitätsaufgaben und acht 1-Minuten-Zeichnungs-Kreativitätsaufgaben (Miller & Tal, 2007). Da eine paarweise orientierte Denkweise die Kreativität fördert (Griskevicius, Cialdini, & Kenrick, 2006), wurden die Teilnehmer gebeten, diese Aufgaben so kreativ wie möglich zu erfüllen, um einen romantischen Partner zu gewinnen. Beispiele für verbale Aufgaben waren das Schreiben von Antworten auf zum Nachdenken anregende Fragen wie "Wie würdest du eine Ehe nach den ersten paar Jahren spannend halten?" "Was erhoffen Sie sich von der Welt in 100 Jahren?" und "Stellen Sie sich vor, alle Wolken hätten wirklich lange Saiten an ihnen hängen – Mehrere hundert Meter lang. Welche Auswirkungen hätte diese Tatsache auf die Natur und die Gesellschaft?" Es gab zwei Arten von Zeichenaufgaben, vier abstrakte (z. B. "Zeichnen Sie ein abstraktes Symbol, Muster oder eine Komposition, die Ihr Glück als Kind darstellt, das eine Lieblingsaktivität macht") und vier gegenständliche (z. B. "Zeichnen Sie im Raum unten ein Tier, das Sie wegen seiner Stärke, Anmut, Geschwindigkeit oder Schönheit bewundern"). Die Antworten jedes Teilnehmers auf jede der 14 Kreativitätsaufgaben wurden unabhängig voneinander von vier Ratern auf einer 1- bis 5-Punkte-Kreativitätsskala bewertet. Die daraus resultierenden zusammengesetzten verbalen Kreativitäts- und Zeichnungs-Kreativitätsmaßnahmen zeigten eine hohe Interrater-Zuverlässigkeit und IC (Cronbachs Alphas = 0,91 bzw. 0,90; Miller & Tal, 2007).

## **Ergebnisse**

### ***Beispiel 1***

IC-Schätzungen wurden mit Cronbachs Alpha berechnet. Der APM-18 zeigte einen moderaten IC ( $b = .71$ ), wobei der eingebettete APM-12 einen etwas niedrigeren Wert lieferte. Obwohl diese internen Konsistenzen niedriger sind als die in Studie 1 berichteten, sind sie immer noch mäßig stark.

Der mittlere APM-18-Wert betrug 10,68,  $SD = 3,25$  (Männlich = 11,07,  $SD$ -Männchen = 3,13; Weiblich = 10,31,  $SD$ weiblich = 3,34), mit einem Bereich von 13 (vier bis 17). Es gab keine Beziehung zwischen APM-18-Score und Alter ( $r = 0,03$ ,  $p = 0,71$ ) oder Geschlecht ( $r = .11$ ,  $p = 0,10$ ). Die Durchschnittswerte für jedes DerIntelligenzmaße sind in Tabelle 2 zu sehen. Aufgrund des signifikanten Geschlechtsunterschieds zwischen den APM-18-Werten in Studie 1 wurden die mittleren Geschlechtsunterschiede bei allen Maßnahmen in dieser Studie durch t-Tests überprüft. Es gab keine

signifikanten Geschlechtsunterschiede für eine der Intelligenzmaßnahmen in dieser Stichprobe, mit Ausnahme eines moderaten männlichen Vorteils auf selbst gemeldeten SAT-Scores ( $t = 3,00$ ,  $p = 0,003$ ). Daher wurden die übrigen Analysen an der vollständigen Stichprobe und nicht an der Geschlechtsprobe durchgeführt.

Wie in Tabelle 3 dargestellt, korrelierten sowohl das APM-18 als auch das eingebettete APM-12 signifikant mit den meisten anderen Messgrößen der Intelligenz und der akademischen Leistung und Eignung, die in dieser Stichprobe verwendet wurden. Insbesondere korrelierten beide APM-Skalen positiv und signifikant am stärksten mit der Shipley Abstraction-Skala und den Selbstanzeigen von SAT-Werten. Dies ist nicht verwunderlich, da das APM als Maß g konzipiert ist, das am leichtesten in Bezug auf abstrakte, analytische Maßnahmen identifiziert werden kann, von denen die Shipley-Abstraktion eine ist, und die SAT eine analytische Teilskala enthält.

### **Beispiel 2**

IC-Schätzungen wurden wiederum mit Cronbachs Alpha berechnet. Wie in Studie 1 und Probe 1 dieser Studie zeigte die APM-18 eine moderate Zuverlässigkeit ( $b = .79$ ), während die eingebettete APM-12 wiederum eine etwas geringere Zuverlässigkeit zeigt ( $b = 0,74$ ). Der mittlere APM-18-Wert betrug 9,53,  $SD = 3,57$  (Mmännlich = 10,29,  $SD$ -Stecker = 3,98; Mweiblich = 9,23,  $SD$ weiblich = 3,36), mit einem Bereich von 1 bis 18. Es gab keine Beziehung zwischen APM-18-Score und Alter ( $r = 0,03$ ,  $p = 0,61$ ), aber es gab zwischen APM-18-Score und Geschlecht ( $r = .13$ ,  $p = 0,04$ ), wobei die Männer wieder etwas höher punkteten ( $t = 2,04$ ,  $p < 0,05$ ). Es gab keine anderen Geschlechtsunterschiede bei den anderen intelligenzisch (für alle  $t$ s 1,20,  $p > 0,05$ ). Tabelle 4 zeigt Durchschnittswerte für die intelligenzischen Maßnahmen und NEO-FFI-Faktoren.

Wie in Tabelle 5 zu sehen ist, standen sowohl die APM-18- als auch die eingebetteten APM-12-Werte signifikant positiv im Zusammenhang mit der verbalen Kreativität ( $r$ s = .36 bzw. .32,  $p < .001$ ), der Kreativität ( $r$ s = .34 bzw. .29,  $p < 001$ ) bzw. dem selbst gemeldeten ACT-Score ( $r$ s = .44 bzw. .45,  $p < .001$ ). Darüber hinaus bezogen sich die APM-Skalen positiv auf die Offenheit ( $r$ s = .26 bzw. .25,  $p < .01$ ) und negativ auf gewissenhafte Fragen (jeweils  $r = .16$ ,  $p < .05$ ), eine Feststellung, die mit früheren Forschungsergebnissen übereinstimmt (Ackerman & Heggestad, 1997; Moutafi, Furnham, & Crump, 2003).

## **DISKUSSION**

Jede Stichprobe in Studie 2 verwendete verschiedene Methoden zur Beurteilung der konvergenten Gültigkeit des APM-18. Beispiel 1 konzentrierte sich auf Beziehungen zwischen dem APM-18 und anderen Standardmaßstäben der Intelligenz und der akademischen Leistung (z. B. verbale Intelligenztests, selbst gemeldete GPA- und SAT-Scores); in der Erwägung, dass Probe 2 die Beziehung zwischen aPM-18, Kreativität, selbst gemeldeten ACT-Scores und Big Five-Persönlichkeitsmerkmalen untersuchte. Beide Studien bestätigten, dass das APM-18 in vorhersehbarer Weise mit diesen Maßnahmen zusammenhängt. Im Allgemeinen zeigten sowohl das APM-18 als auch das eingebettete APM-12 das gleiche Muster von Korrelationen mit den anderen in diesen Studien verwendeten Messgrößen. Der höhere IC des APM-18 deutet jedoch darauf hin, dass es besser sein könnte, individuelle Variationen in g.

## **SCHLUSSFOLGERUNG**

Jedes der 18 Elemente, die in diesem neuen APM-18-Test verwendet wurden, wurde ausgewählt, um die fortschreitende Schwierigkeit sowohl der Langform (APM-36) als auch der Kurzform (APM-12) beizubehalten. Es überrascht nicht, dass die Zuverlässigkeit des APM-18 zwar niedriger war als die des APM-36, aber höher war als die des von Arthur and Day (1994) entwickelten APM-12. Darüber hinaus sind die Korrelationsmuster mit anderen Intelligenzmaßstäben praktisch identisch mit dem APM-12, das in früheren Studien gezeigt hat, dass es die APM-36-Ergebnisse imitiert (Arthur & Day, 1994; Arthur et al., 1999). In Kombination mit einer durchschnittlichen Verabreichungszeit von 17,53 min (maximal 25 min) deuten diese Ergebnisse darauf hin, dass das APM-18 als Kompromiss für Forscher funktionieren könnte,

die in recht kurzer Zeit ein ziemlich genaues Maß an allgemeiner Intelligenz wünschen. Die Kreuzvalidierung in den drei hier gemeldeten Beispielen ist ein erster Versuch, normative Daten für das APM-18 zu sammeln. Unsere Ergebnisse dürfen nur auf andere College-Studenten verallgemeinert werden. Die kurze Administrationszeit des APM-18, die hohe IC, die angemessene Gültigkeit und die einfache Verwaltung durch Papier und Bleistift in großen Klassenzimmern machen es jedoch ideal für verhaltenswissenschaftliche Studien, bei denen Forscher eine relativ schnelle, genaue Intelligenzbewertung als Teil einer größeren Fragebogenbatterie wünschen.

## **TRANSLATED VERSION: PORTUGUESE**

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

## **VERSÃO TRADUZIDA: PORTUGUÊS**

Aqui está uma tradução aproximada das ideias acima apresentadas. Isto foi feito para dar uma compreensão geral das ideias apresentadas no documento. Por favor, desculpe todos os erros gramaticais e não responsabilize os autores originais responsáveis por estes erros.

## **INTRODUÇÃO**

O Teste de Matrizes Progressivas de Ravens, desenvolvido pela Raven (1941) como uma medida de inteligência geral (g), passou por muitas revisões, desde versões coloridas para crianças até as matrizes padrão e avançada para adultos de diferentes níveis cognitivos. A versão mais recente publicada é a Ravens Advanced Progressive Matrices (APM; Raven, Raven, & Court, 1993), que foi desenvolvido para populações adultas de maior capacidade (ou seja, nível universitário e superior). Este teste é construído com 36 itens de dificuldade crescente divididos em três conjuntos de 12 itens; em cada item, o examinado é solicitado a completar um padrão visual escolhendo uma das oito soluções possíveis.

Devido ao seu formato não verbal, a APM é supostamente uma medida culturalmente justa e imparcial de inteligência fluida (Cattell, 1963), habilidade educativa (J. Raven et al., 1993), ou, como se referirá a ela, inteligência geral (g; Spearman, 1927), e tem se mostrado especialmente útil em situações onde o inglês não é a língua primária de um indivíduo. Como tal, as Matrizes Progressivas Padrão e Avançada têm sido amplamente utilizadas em muitas configurações aplicadas nos Estados Unidos (por exemplo, Ackerman, 1992) e em muitas culturas (Owen, 1992; J. C. Raven, 2000; Rushton, Cvorovic, & Bons, 2007). No entanto, os aspectos positivos deste teste são marcados pelo seu longo tempo de administração (40-60 min), dificultando o uso em pesquisas multivariadas com tempo ou ambientes em sala de aula.

Em resposta a essas várias limitações, Arthur e Day (1994) desenvolveram uma forma curta de 12 itens da APM (que chamamos de APM-12), com um tempo de administração de 15 minutos. Vários estudos mostraram que esta forma de 12 itens mostra propriedades psicométricas aceitáveis (por exemplo, alfa de Cronbach, confiabilidade de teste-reteste, validade convergente; ver Arthur, Tubre, Paul, & Sanchez-Ku, 1999, para revisão). No entanto, esta forma curta mostra consistência interna relativamente baixa e variável (IC). Por exemplo, os alfas de Cronbach variam de 0,58 a .66 para forma curta e 0,72 a .73 para os 12 itens de forma curta extraídos da versão completa de 36 itens (Arthur & Day, 1994).

Mais recentemente, Hamel e Schmittmann (2006) argumentaram que o APM completo de 36 itens pode ser administrado como um teste de velocidade de 20 minutos. As pontuações nesta forma acelerada do APM mostram fortes correlações com escores em versões temporizadas mais lentas (40 min,  $r = 0,74$ ) e versões não cronometradas ( $r = 0,75$ ) do APM. No entanto, esses autores não relataram o IC da Escala

de Teste de Velocidade. Também suspeitamos que dar aos adultos típicos apenas 20 minutos para completar 36 problemas abstratos muito desafiadores pode impor estresse indevido.

O objetivo do presente estudo foi desenvolver uma versão de forma média do APM que resultou em IC superior à versão de 12 itens (APM-12), mas menor tempo de administração do que o APM (APM-36) completo de 36 itens — uma combinação de recursos que podem ser úteis para situações de teste de tempo e teste de massa. Aqui, relatamos o desenvolvimento e a validade de construção desta escala de 18 itens.

## **ESTUDO 1: CONSTRUÇÃO EM ESCALA E VALIDADE DE CONSTRUÇÃO**

### **Método**

#### ***Participantes***

Um total de 633 estudantes (198 homens, 435 mulheres) de três universidades do sudoeste participaram deste estudo como requisito parcial para o crédito experimental do curso. A média de idade para os participantes foi de 20,92, SD = 4,07 (Mmale = 20,85, SDmale = 3,90; Mfemale = 20,96, SDfemale = 4,15). As idades variaram de 17 a 58 anos (homem = 18-41 anos, feminino = 17-58).

### **Medidas**

#### ***Os Ravens Advanced Progressive Matrices 18-Item Short Form (APM-18)***

Esta versão de forma curta de 18 itens do APM é impressa em um formato de livreto em papel branco de 8 1/2" por 11", com cada item de teste impresso em uma página separada. As quatro primeiras páginas do folheto de teste contêm três itens de exemplo (Itens de prática 1, 5 e 9 da APM-36) para explicar a tarefa.

Os 18 itens reais de teste foram derivados adicionando seis itens da versão mais longa de 36 itens (J. Raven et al., 1993) a Arthur and Day's (1994) publicado versão de 12 itens. Arthur e Day utilizaram os itens 1, 4, 8, 11, 15, 18, 21, 23, 25, 30, 31 e 35 do APM de 36 itens com base em um conjunto de três regras de decisão, que podem ser resumidas como (a) dividindo o APM em 12, três seções com base na dificuldade; b Levar o item com a maior correlação item-total para cada seção; e (c) no caso de empate, incluindo o item que resultou na maior queda no IC se for excluído do teste completo. Seguindo essas mesmas regras, adicionamos mais seis itens de dificuldade crescente — dois que foram fáceis (96% e 75% dos exames da amostra normativa responderam corretamente), dois moderados (50% e 48% dos exames da amostra normativa responderam corretamente) e dois que foram difíceis (37% e 32% dos exames da amostra normativa responderam corretamente). Esses itens (2, 20, 22, 24, 34 e 32) foram integrados de dificuldade para imitar sua ordem de apresentação no APM original.

### ***Procedimento***

O novo teste APM-18 foi ministrado em ambientes de sala de aula com vários exames de cada vez. Isso foi feito porque esse teste foi desenvolvido como uma medida de g que poderia ser utilizado em ambientes como salas de aula, onde há prazos nas sessões de pesquisa. Em uma subsample (n = 175), foram realizados testes sem restrições de tempo, mas com tempos de conclusão registrados, para determinar o tempo médio necessário para a conclusão. As outras duas subsamples (n = 232 e n = 226) foram constrangidas a terminar o teste dentro de 25 minutos, sem tempo de conclusão individual registrado.

### ***Análises***

Todas as análises estatísticas foram realizadas utilizando-se a versão SAS 8.2 (Instituto SAS, 1999). As correlações alfa e bivariada de Cronbach foram computadas por meio do procedimento PROC CORR. Os testes para diferenças médias entre os sexos foram calculados por meio de procedimentos de teste T



(PROC TTEST). Os modelos lineares gerais hierárquicos (GLMs) foram testados utilizando-se o PROC GLM.

## Resultados

As estimativas de IC foram calculadas usando o alfa de Cronbach. O IC da escala APM-18 produziu confiabilidade moderada ( $\alpha = 0,79$ ). Este alfa é inferior aos relatórios normativos de IC para o APM-36 ( $\alpha = .84$ ; Forbes, 1964), mas maior do que os do APM-12 (variando de  $\alpha = 0,58-66$ ; ver Arthur et al., 1999). Além disso, o alfa do APM-18 era maior que o do APM-12 incorporado ( $\alpha = .73$ ). A Tabela 1 mostra os resultados de cada um dos itens APM-18, com relação às suas correlações totais de itens, dificuldades de itens e escala  $\alpha$  da escala geral se o item for excluído. Como visto, a exclusão de qualquer item reduz a confiabilidade geral da escala, sugerindo que todos os itens devem ser retidos.

A pontuação média APM-18 foi de 9,73, SD = 3,59 (Mmasculino = 10,43, SD masculino = 3,52; Mfeminino = 9,41, SD feminino = 3,59), com intervalo de 18. Para a subsamplena em que os tempos de conclusão foram registrados ( $n = 175$ ), o tempo médio de conclusão do teste foi de 17,5 min (SD = 4,67), com intervalo de 7 a 25 min; 21% dos participantes demoraram mais de 20 minutos, mas ninguém demorou mais de 25 minutos. Nessa subsampleidade, houve uma relação positiva significativa entre o tempo que levou para os participantes fazerem o teste e a pontuação APM-18 ( $r = 0,41$ ,  $p < .001$ ), mas não houve relação entre idade e escore APM-18, ou idade e tempo necessários para concluir o teste (para cada  $r \leq 0,3$ ,  $p \geq .71$ ). No entanto, para a amostra completa ( $N = 633$ ), os participantes mais jovens obtiveram pontuação um pouco maior (idade e Escores APM-18 correlacionados  $r = -.15$ ,  $p < .001$ ), e os machos obtiveram pontuação um pouco maior — sexo (feminino = 0, masculino = 1) e APM-18 correlacionados  $r = 0,13$ ,  $p < .001$ .

GIMs hierárquicos foram testados para explorar se as diferenças aparentes nos escores APM-18 masculino e feminino poderiam ter sido indiretamente atribuíveis à relação entre a idade e os escores APM-18. Este modelo definiu o escore APM-18 como variável critério, com as variáveis preditores ordenadas sendo idade e, em seguida, sexo. O modelo hierárquico foi projetado para permitir que a idade absorva o máximo de variância possível, com o sexo inserido no modelo apenas depois. Utilizando este modelo, ambos os GLMs indicaram um efeito significativo para a idade ( $F = 15,39$ ,  $p < .001$ ) e, em seguida, também para o sexo após a idade tinha sido estatisticamente controlado ( $F = 10,66$ ,  $p = 0,001$ ).

## Discussão

Os resultados aqui apresentados sugerem que o APM-18 pode servir como um compromisso útil entre a menor confiabilidade APM-12 e a APM-36 muito mais longa. Os GLM hierárquicos identificam idade e sexo como preditores significativos dos escores APM-18, com indivíduos mais jovens e homens geralmente pontuando mais alto. Esses resultados são consistentes com muitos estudos anteriores que analisam a inteligência geral (por exemplo, Jackson & Rushton, 2006). Os resultados do Estudo 1, no entanto, não testam a validade convergente dessa escala em relação a outras medidas de inteligência. O estudo 2 foi projetado para fazer isso.

## ESTUDO 2: VALIDADE CONVERGENTE

O estudo 2 foi realizado para avaliar a validade convergente da APM-18 com outras medidas de inteligência, realização acadêmica e personalidade. Para tanto, testamos duas subamostras separadas ( $n = 193$  e  $229$ ) retiradas do Estudo 1, cada uma das quais utilizaram diferentes medidas de critério. Na Amostra 1, foram utilizadas duas medidas amplamente utilizadas de inteligência adulta: a Escala de Vocabulário mill-hill - Conjuntos de Múltipla Escolha A & B (MHV-MC; J. Raven, Raven, & Court,

1997), desenvolvido para ser usado em conjunto com o APM-36 como medida de capacidade reprodutiva, ou seja, a capacidade de armazenar e recuperar informações (J. C. Raven, 1989); e o Shipley Institute of Living Scale (SILS; Zachary, 1986), que é um teste de inteligência independente composto por duas subescalas — Vocabulário, que testa inteligência cristalizada, e Abstração, que testa inteligência fluida. Além disso, examinamos o desempenho acadêmico através da média de pontos de grau autorrenotada (GPA) e dos escores do Teste de Aptidão Escolar (SAT). Na Amostra 2, examinamos correlações entre as escores APM-18 e as dimensões de personalidade do Big Five avaliadas com a Escala neo cinco fatores de inventário (NEO-FFI) (Costa & McCrae, 1992), e criatividade verbal e de desenho (Miller & Tal, 2007). Além disso, os escores do ACT foram coletados nesta segunda amostra.

## **Método**

### ***Participantes***

A amostra 1 foi composta por 193 estudantes (94 homens, 99 mulheres) de um curso de psicologia introdutória na Universidade do Arizona. A média de idade dos participantes foi de 19,11, SD = 1,62 (Mmasculino = 19,23, SD masculino = 1,07; Mfeminino = 19,01, SD feminino = 2,00). Devido ao tempo necessário para administrar o APM-18, o Shipley e o Mill-Hill, 10 participantes não completaram o Teste Mill-Hill. Pedimos aos participantes que registrassem suas pontuações de SAT, ACT e GPA apenas se estivessem certos deles; devido a essa restrição, muitas dessas pontuações também estavam faltando.

A amostra 2 foi composta por 229 estudantes (65 homens, 164 mulheres) de diversos cursos de graduação da Universidade do Novo México. A média de idade dos participantes foi de 20,19, SD = 3,43 (Mmale = 21,05, SDmasculino = 5,01; Mfeminino = 19,85, SDfemale = 2,48). Mais uma vez, pedimos aos participantes que registrassem suas pontuações do ACT apenas se estivessem certos deles, deixando-nos com pontuações act para apenas 129 participantes.

### **Medidas**

#### ***APM-18***

O APM-18 consistiu nos mesmos itens identificados no Estudo 1. Na Amostra 1, o formulário foi apresentado primeiro em uma série de medidas que examinam a inteligência adulta. Na Amostra 2, foi apresentado no meio de um pacote de questionários sobre personalidade, criatividade, comportamento sexual e inteligência.

#### ***O SILS***

O SILS (Zachary, 1986) é um temporizado (10 min por subescala), medida de auto-relato de 60 itens que examina tanto a inteligência verbal (40 itens) quanto a inteligência abstrata (20 itens). O teste é considerado adequado para indivíduos de língua inglesa média de 14 a 14 anos adultos, que são tomadores de teste motivados. As validades e normas publicadas no manual foram retiradas de uma amostra de 322 recrutas do exército. As capacidades divididas para cada subescala são relatadas como 0,87 para Vocabulário, 0,89 para Abstração e 0,92 para a pontuação total.

#### ***O MHV-MC***

O MHV-MC (J. Raven et al., 1997) é um teste de vocabulário auto-administrado de 68 itens projetado para complementar o APM-36. Considerando que o APM visava medir a capacidade de um indivíduo de resolver novos problemas e pensar de novas formas (ou seja, inteligência fluida), o Mill-Hill visava medir a capacidade de um indivíduo de recordar informações aprendidas (ou seja, inteligência cristalizada). Nessa medida, indica realização educacional, formação cultural e familiaridade com a linguagem do teste. O Mill-Hill normalmente mostra mais de 0,90 reabilizações de teste que variam entre 0,87 e 0,95 (Raven et al., 1997).

### ***Desempenho acadêmico***

O desempenho acadêmico foi medido por GPAs autorrenotados e escores de SAT na Amostra 1. Foram solicitados aos participantes da amostra 2 as pontuações do SAT e act. Uma variedade de estudos

identificou correlações moderadas a fortes entre essas medidas de realização acadêmica e aptidão, e uma variedade de outros traços, incluindo inteligência, personalidade e psicopatologia (Barton, Dielman, & Cattell, 1971; Brown, 1994; Dyer, 1987; Mouw & Khanna, 1993).

### ***NEO-FFI***

O NEO-FFI (Costa & McCrae, 1992) é a medida mais utilizada em pesquisas sobre o modelo de personalidade de Cinco Fatores. É uma versão encurtada do Inventário de Personalidade NEO Revisado de 240 itens (NEO-PI-R; Costa & McCrae, 1992), composta por 60 itens que medem cinco fatores globais de personalidade (12 itens por fator): Abertura à Experiência, Consciência, Extraversão, Concordação e Neuroticismo. Em nossa versão, os participantes avaliaram o grau de concordância com declarações sobre suas personalidades e propensões comportamentais em uma escala de 5 pontos que varia de -2 (discordo fortemente) a 0 (neutro) a +2 (fortemente concordo). Esta escala tem mostrado ic forte, com os alfas de Cronbach variando entre 0,74 a 0,89, para cada fator, e validade transcultural consistente (McCrae & Costa, 1997).

### **Tarefas verbais e de desenho de criatividade**

Os participantes completaram seis tarefas de criatividade verbal de 2 minutos e oito tarefas de criatividade de desenho de 1 min (Miller & Tal, 2007). Como uma mentalidade orientada para acasalamento promove a criatividade (Griskevicius, Cialdini, & Kenrick, 2006), os participantes foram convidados a completar essas tarefas da forma mais criativa possível com a intenção de atrair um parceiro romântico. Exemplos de tarefas verbais incluíam escrever respostas para perguntas instigantes, como "Como você manteria um casamento emocionante após os primeiros anos?" "Como você espera que o mundo seja em 100 anos?" e "Imagine que todas as nuvens tinham cordas muito longas penduradas neles — cordas de centenas de metros de comprimento. Quais seriam as implicações desse fato para a natureza e a sociedade?" Havia dois tipos de tarefas de desenho, quatro resumos (por exemplo, "Por favor, desenhe um símbolo abstrato, padrão ou composição que represente sua felicidade quando criança fazendo uma atividade favorita") e quatro representativos (por exemplo, "No espaço abaixo, por favor, desenhe um animal que você admira por sua força, graça, velocidade ou beleza"). As respostas de cada participante a cada uma das 14 tarefas de criatividade foram pontuadas independentemente por quatro avaliadores em uma escala de criatividade de 1 a 5 pontos. As medidas de criatividade e criatividade compostas resultantes apresentaram alta confiabilidade interrater e IC (alfas de Cronbach = 0,91 e 0,90, respectivamente; Miller & Tal, 2007).

## **Resultados**

### ***Amostra 1***

As estimativas de IC foram calculadas usando o alfa de Cronbach. O APM-18 apresentou IC moderado ( $\alpha = 0,71$ ), com o APM-12 incorporado produzindo um valor ligeiramente menor ( $\alpha = 0,63$ ). Embora essas consistências internas sejam inferiores às relatadas no Estudo 1, elas ainda são moderadas em força.

A pontuação média APM-18 foi de 10,68, SD = 3,25 (Mmasculino = 11,07, SD masculino = 3,13; Mfeminino = 10,31, SD feminino = 3,34), com intervalo de 13 (quatro a 17). Não houve relação entre escore APM-18 e idade ( $r = 0,03$ ,  $p = 0,71$ ) ou sexo ( $r = 0,11$ ,  $p = .10$ ). Pontuações médias para cada uma das medidas de inteligência podem ser vistas na Tabela 2. Devido à diferença de sexo

significativa entre os escores de APM-18 no Estudo 1, as diferenças sexuais médias em todas as medidas deste estudo foram verificadas via t testes. Não houve diferenças sexuais significativas para nenhuma das medidas de inteligência nesta amostra, exceto para uma vantagem masculina moderada sobre os escores de SAT autorrenotados ( $t = -3,00$ ,  $p = 0,003$ ). Portanto, as demais análises foram realizadas na amostra completa e não por sexo.

Como visto na Tabela 3, tanto o APM-18 quanto o APM-12 incorporado correlacionam-se significativamente com a maioria das outras medidas de inteligência e desempenho acadêmico e aptidão utilizadas nesta amostra. Especificamente, as escalas APM correlacionam-se positiva e significativamente mais fortemente com a escala de Abstração shIPLEY e os escores sat de auto-relato. Isso não é surpreendente, pois o APM foi projetado para ser uma medida de g, que pode ser mais facilmente identificada em relação a medidas abstratas e analíticas, das quais a abstração shIPLEY é uma, e o SAT contém uma subescala analítica.

### **Amostra 2**

As estimativas de IC foram novamente calculadas usando o alfa de Cronbach. Como no Estudo 1 e Amostra 1 deste estudo, o APM-18 mostrou confiabilidade moderada ( $\alpha = 0,79$ ), enquanto o APM-12 incorporado novamente mostra confiabilidade ligeiramente menor ( $\alpha = 0,74$ ). A pontuação média APM-18 foi de 9,53, SD = 3,57 (Mmasculino = 10,29, SD masculino = 3,98; Mfeminino = 9,23, SD feminino = 3,36), com intervalo de 1 a 18. Não houve relação entre escore APM-18 e idade ( $r = -0,3$ ,  $p = 0,61$ ), mas houve entre escore APM-18 e sexo ( $r = 0,13$ ,  $p = 0,04$ ), com os machos novamente pontuando ligeiramente mais alto ( $t = -2,04$ ,  $p < .05$ ). Não houve outras diferenças sexuais nas outras medidas de inteligência (para todos os ts  $\geq -1,20$ ,  $p > .05$ ). A Tabela 4 mostra pontuações médias para as medidas de inteligência e fatores NEO-FFI.

Como visto na Tabela 5, tanto os escores APM-18 quanto os APM-12 incorporados foram significativamente relacionados positivamente à criatividade verbal ( $r_s = 0,36$  e  $.32$ ,  $p < .001$ , respectivamente), criatividade de desenho ( $r_s = 0,34$  e  $.29$ ,  $p < .001$ , respectivamente) e escore ACT autorrenotado ( $r_s = 0,44$  e  $.45$ ,  $p < .001$ , respectivamente). Além disso, as escalas APM foram positivamente relacionadas à Abertura ( $r_s = 0,26$  e  $.25$ ,  $p < .01$ , respectivamente) e negativamente à Consciência (para cada,  $r = -.16$ ,  $p < .05$ ), um achado consistente com pesquisas anteriores (Ackerman & Heggestad, 1997; Moutafi, Furnham, & Crump, 2003).

## **DISCUSSÃO**

Cada amostra do Estudo 2 utilizou diferentes métodos de avaliação da validade convergente do APM-18. A amostra 1 concentrou-se nas relações entre o APM-18 e outras medidas padrão de inteligência e realização acadêmica (por exemplo, testes de inteligência verbal, GPA autorrelato e escores de SAT); que a Amostra 2 examinou a relação entre o APM-18, criatividade, pontuações de ACT autorrenotadas e traços de personalidade do Big Five. Ambos os estudos confirmaram que o APM-18 está relacionado a essas medidas de forma previsível. De um modo geral, tanto o APM-18 quanto o APM-12 incorporado apresentaram o mesmo padrão de correlações com as demais medidas utilizadas nesses estudos. No entanto, o IC mais alto do APM-18 sugere que ele pode ser melhor na detecção de variação individual em g.

## **CONCLUSÃO**

Cada um dos 18 itens utilizados neste novo teste APM-18 foi escolhido para manter a dificuldade progressiva tanto da forma longa (APM-36) quanto da forma curta (APM-12). Sem surpresa, embora a confiabilidade do APM-18 tenha sido inferior à do APM-36, foi maior que a da APM-12 desenvolvida por Arthur e Day (1994). Além disso, os padrões de correlação com outras medidas de inteligência são praticamente idênticos ao APM-12, que, em estudos anteriores, tem mostrado imitar os resultados da APM-36 (Arthur & Day, 1994; Arthur et al., 1999). Combinados com um tempo médio de administração de 17,53 min (25 min no máximo), esses achados sugerem que o APM-18 pode funcionar bem como um compromisso para pesquisadores que querem uma medida bastante precisa de inteligência geral em um período bastante curto de tempo. A validação cruzada nas três amostras aqui relatadas é uma tentativa inicial de coleta de dados normativos para o APM-18. Nossos resultados podem generalizar apenas para outros estudantes universitários. No entanto, o curto tempo de administração do APM-18, o alto IC, a validade razoável e a facilidade de administração por papel e lápis em grandes ambientes de sala de aula universitários tornam-no ideal para estudos de ciência comportamental onde os pesquisadores querem uma pontuação de inteligência razoavelmente rápida e precisa como parte de uma bateria de questionário maior.