THE DEVELOPMENT AND VALIDATION OF A RELIABLE ALTERNATE FORM FOR RAVEN'S STANDARD PROGRESSIVE MATRICES

Nancy E. Johnson
Dennis P. Saccuzzo
Tracey L. Guertin
San Diego State University

The development of an alternate form for Raven's Standard Progressive Matrices (SPM) Test is described. Items for each of the original 60 items of the SPM were developed to be comparable to the corresponding original items in terms of the underlying strategy and difficulty. An alternate form reliability analysis on a diverse group of 449 children (77 African-American, 122 Asian, 54 Filipino, 156 Latino/Hispanic, 38 white, and 2 other) showed an alternate form reliability coefficient of .90. Kuder-Richardson reliabilities of the newly developed alternate and the SPM were identical at .94. A Kolmogorov-Smirnov two-sample test, moreover, revealed no significant differences in central tendency, dispersion, and skewness for distributions of individual item difficulties. In addition, the two tests showed comparable predictive validity coefficients. The alternate form resolves one limitation of the SPM and could provide widespread utility as a research tool.

The Raven Progressive Matrices Tests, including the Standard (SPM) and Advanced (APM) versions, are among the most widely used and referenced of the standardized psychometric tests of ability (Carpenter, Just, & Shell, 1990; Kaplan & Saccuzzo, 1993; Raven, 1989). The SPM, for example, is extensively used in clinical, neuropsychological, educational, and research settings (Carpenter et al., 1990; Cherkes-Julkowski, Stolzenberg, & Segal, 1990; Larson, Alderton, & Kaupp, 1991). Of the psychometric ability tests, the SPM and related versions are among the best measures of Spearman's (1927) g factor (Marshalek, Lohman, & Snow, 1983; Snow, Kyllonen, & Marshalek, 1984). Indeed, perhaps more is known about what the SPM and related versions measure than about any other ability test. Positron Emission Tomography (PET) Scan studies have revealed that the entire brain—especially the right cerebral hemisphere, left temporal lobe, and left frontal lobe—is used in solving APM problems (Haier et al., 1988). Theoretical reviews have pointed to the SPM and related versions as measures of the "ability to reason and solve problems involving new information, without relying extensively on an explicit base of declarative knowledge" (Carpenter et al., 1990, p. 404).

Until recently, the application of the SPM was limited due to the absence of adequate norms and of an alternate form. Now that an extensive and
Johnson, Saccuzzo, and Guertin

relatively current set of U.S. and worldwide norms is available (Raven, 1986), the first of these limitations has finally been overcome. In addition, to facilitate the research use of the SPM and permit matching across various forms, Jensen, Saccuzzo, and Larson (1988) developed a set of standards for equating the SPM and APM. To date, however, the absence of an alternate form remains a limitation.

An alternate form is needed for a variety of reasons. One such reason would be the measurement of an intervention effect. In neuropsychological research, for example, it is often desirable to evaluate the effects of an intervention within a time frame that is shorter than the optimal duration for a test-retest paradigm. In such cases, an alternate form would be extremely useful to evaluate intervention effects while minimizing the effects of practice. In our own research, which involves group testing, an alternate is needed as a replacement for invalid test administrations. Thus, an alternate form has potential utility as a research tool in neuropsychology, education, intellectual assessment and other areas of psychological inquiry.

In the present study, we report on the development, validity, and reliability of an alternate form of the SPM called the San Diego Test of Reasoning Ability (SANTRA) (Johnson et al., 1993). In developing this alternate form, we began with the 12 item rules as explicated by Jacobs and Vandeventer (1972), such as "shading" (i.e., progressive change in shading), "number series" (constant increase in items across cells), and "mirror image" (i.e., figure moves as if lifted up and replaced face down).

For each of the 60 items in the SPM, we determined the item rule(s) underlying solution of the matrix. Two researchers discussed the rule(s) and attempted to develop an entirely new item that was based on the same item rule(s) and designed to be of comparable difficulty and complexity. The item was submitted to a third researcher who independently checked the item in terms of the underlying rule(s). If the third person disagreed, there was further discussion and a new item was submitted. This process continued until there was complete agreement. Our initial psychometric evaluation of the alternate form is based on a cross section of children from a large, diverse school district.

Method

Participants

The participants were 449 second-, fifth-, and seventh-grade students. Permission was obtained from the school district to select entire classes that were believed to reflect the San Diego City School District of 123,000 children in terms of diversity of the children. In this sample, 77 were African-American, 122 Asian, 54 Filipino, 156 Latino/Hispanic, 38 white, and 2 other. Of these 214 were boys, 233 were girls, and for 2 gender data were missing. The mean age of the children was 11 years (range = 6 years, 8 months–13 years, 10 months).

Procedure

Parents of the children provided written informed consent for participation in the study. Each child was administered both the SPM and the alternate (SANTRA). Approximately half (216) of the children were randomly assigned to a SPM-alternate form order of test administration and half to an alternate-SPM order. Upon finishing the first test, each participant turned in the test protocol and was given the form not yet taken. Testing was untimed; children were told to work as quickly and accurately as possible but that there was no time limit. Total testing time ranged from approximately 1½ to 2½ hours.

Results

The overall correlation (Pearson r) between the SPM and the newly developed alternate form (i.e., alternate form reliability coefficient) was .90 (p < .01). The Kuder Richardson (KR-20) reliabilities for the SPM and alternate forms were .9447 and .9442, respectively. Thus, the two forms were highly comparable.

To examine item difficulty for the two forms, we calculated the percentage of children who obtained correct responses on each of the 60 items comprising each of the two forms (see Table 1). Distributions of item difficulty were compared using a Kolmogorov-Smirnov two-sample test. As Siegel (1956) noted, this nonparametric test "is sensitive to any kind of difference in the
distributions from which the two samples [of items] were drawn—differences in location (central tendency), in dispersion, in skewness, etc."

(p. 127). The results showed no statistically significant difference (p > .06) between the two distributions of individual item difficulties, thus supporting the comparability of the two forms.

Mean number correct was 36.10 (SD = 11.52) for the SPM and 35.46 (SD = 11.93) for the alternate. The slight difference in these means was statistically significant, t(448) = -2.56, p < .01. Table 2 contains the distribution of SPM minus SANTRA discrepancy scores. This frequency distribution revealed a negative skew (skewness = -.11) consistent with the slight but significant difference in the overall test means. Thus, although the SANTRA can be considered an alternate form, the SANTRA was slightly more difficult than the SPM.

For a subset of the sample (n = 32), achievement test scores were available. For these children, the two forms were correlated (Pearson r) with the Comprehensive Test of Basic Skills (CTBS; 1982) Total Language, Total Reading, and Total Math. Table 3 shows the intercorrelation matrix. Inspection of Table 3 reveals that both forms were approximately equivalent in the correlation with each of the three achievement measures, with a range of correlation coefficients of .42 to .56.

| Change score | Frequency | Percent |
|--------------|-----------|---------|
| < -10        | 7         | 1.6     |
| -10          | 1         | .2      |
| -9           | 5         | 1.1     |
| -8           | 6         | 1.3     |
| -7           | 4         | .9      |
| -6           | 7         | 1.6     |
| -5           | 20        | 4.5     |
| -4           | 22        | 4.9     |
| -3           | 37        | 8.2     |
| -2           | 43        | 9.6     |
| -1           | 36        | 8.0     |
| 0            | 36        | 8.0     |
| 1            | 46        | 10.2    |
| 2            | 33        | 7.3     |
| 3            | 34        | 7.6     |
| 4            | 26        | 5.8     |
| 5            | 20        | 4.5     |
| 6            | 21        | 4.7     |
| 7            | 10        | 2.2     |
| 8            | 11        | 2.4     |
| 9            | 6         | 1.3     |
| 10           | 3         | .7      |
| > 10         | 15        | 3.3     |

Note. SPM = Standard Progressive Matrices; SANTRA = San Diego Test of Reasoning Ability.

Table 1

| Item Difficulties Expressed as Percentage Correct |
|--------------------------------------------------|
| Item    | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 | A11 | A12 |
| SPM     | 97.3 | 97.8 | 98.4 | 98.4 | 97.6 | 98.0 | 95.7 | 98.0 | 91.8 | 84.4 | 84.1 | 60.1 | 42.5 |
| SANTRA  | 97.8 | 97.8 | 97.6 | 93.8 | 89.8 | 91.1 | 49.0 | 99.0 | 81.1 | 89.8 | 72.6 | 80.0 | 57.5 |
| Item    | B1 | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B9 | B10 | B11 | B12 |
| SPM     | 97.1 | 95.8 | 88.4 | 80.4 | 75.7 | 73.1 | 56.6 | 67.0 | 70.4 | 77.1 | 68.6 | 38.3 |
| SANTRA  | 96.4 | 85.7 | 76.4 | 70.8 | 84.4 | 94.4 | 71.1 | 66.4 | 76.4 | 78.8 | 70.6 | 35.2 |
| Item    | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 |
| SPM     | 86.2 | 82.2 | 78.0 | 65.9 | 76.6 | 57.2 | 63.5 | 45.4 | 66.8 | 35.0 | 27.2 | 3.6 |
| SANTRA  | 86.0 | 65.3 | 71.0 | 45.1 | 78.8 | 69.7 | 72.4 | 59.2 | 55.7 | 40.5 | 60.6 | 11.6 |
| Item    | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | D11 | D12 |
| SPM     | 89.5 | 81.7 | 76.8 | 70.6 | 80.2 | 66.4 | 55.9 | 58.4 | 47.2 | 52.3 | 21.2 | 8.0 |
| SANTRA  | 87.1 | 81.3 | 57.2 | 67.9 | 67.5 | 54.1 | 59.9 | 65.5 | 43.2 | 38.3 | 24.7 | 47.7 |
| Item    | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | E10 | E11 | E12 |
| SPM     | 60.1 | 47.2 | 39.6 | 24.7 | 29.0 | 22.9 | 20.9 | 14.5 | 11.4 | 5.1 | 4.0 | 8.0 |
| SANTRA  | 17.8 | 19.4 | 12.7 | 32.1 | 32.3 | 20.0 | 19.4 | 10.5 | 22.9 | 8.2 | 22.5 | 5.8 |

Note. SPM = Standard Progressive Matrices; SANTRA = San Diego Test of Reasoning Ability. A-E refer to each of the 5 subsets of 12 items.
The SANTRA exhibited considerable potential as a replacement for the SPM. Moreover, as the SANTRA, proved to be slightly more difficult than the SPM. Thus, the SANTRA can best be described as an alternate, but not an item-by-item parallel form. Clearly, the SANTRA is for research purposes and is not meant for clinical purposes or as a replacement for the SPM. Moreover, as the present psychometric data are based on children (ages 6 years, 8 months–13 years, 10 months) from one geographic location, more work is needed to determine whether the present positive findings generalize to adults and to other populations. In addition, more work is needed to determine whether the SANTRA can be used to predict other criterion variables.¹

### Discussion

In addition to substantial alternate form reliability, the SANTRA showed predictive validity comparable to that for the SPM for achievement test data. The highest correlate for both forms was with Total Math (r = .56 for both forms). The predictive validity coefficients were also substantial for both Total Language and Total Reading. These correlations between verbal measures of achievement and nonverbal measures of intelligence can best be explained in terms of the SPM’s correlation with Spearman’s g factor and Haier et al.’s PET studies. First, to the extent that any verbal measure is related to g, it should share common variance with any other measure of g. Moreover, given the involvement of the left temporal lobes in solving APM problems, it is likely that subjects use a verbal strategy when solving SPM-type stimuli. Hence, verbal skills are utilized in the solution of the types of problems represented by the SPM and the alternate form, the SANTRA.

The SANTRA exhibited considerable potential as a highly reliable alternate form for the SPM. The SANTRA proved to be slightly more difficult than the SPM. Thus, the SANTRA can best be described as an alternate, but not an item-by-item parallel form. Clearly, the SANTRA is for research purposes and is not meant for clinical purposes or as a replacement for the SPM. Moreover, as the present psychometric data are based on children (ages 6 years, 8 months–13 years, 10 months) from one geographic location, more work is needed to determine whether the present positive findings

¹The SANTRA is available upon request for the cost of materials to qualified psychological researchers who agree to use the alternate form for research purposes only. (Available from N. E. Johnson, Ph.D. or D. P. Saccuzzo, Ph.D., 6363 Alvarado Court, Suite 103, San Diego, CA 92120-4913).

### References

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

Cherkes-Julkowski, M., Stolzenberg, J., & Segal, L. (1990). Prompted cognitive testing as a diagnostic compensation for attentional deficits: The Raven Standard Progressive Matrices and attention deficit disorder. *Learning Disabilities, 2*, 1-7.

*Comprehensive Test of Basic Skills*. (1982). Monterey, CA: McGraw-Hill.

Haier, R. J., Siegel, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., Browning, H. L., & Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence, 12*, 199-217.

Jacobs, P. I., & Vandeventer, M. (1972). Evaluating the teaching of intelligence. *Educational and Psychological Measurement, 32*, 235-248.

Jensen, A. R., Saccuzzo, D. P., & Larson, G. E. (1988). Equating the standard and advanced forms of the Raven Progressive Matrices. *Educational and Psychological Measurement, 48*, 1091-1095.

Johnson, N. E., Saccuzzo, D. P., Larson, G. E., Guertin, T. L., Christianson, L., & Longley, S. (1993). *The San Diego Test of Reasoning Ability (SANTRA)*. San Diego: Author.

Kaplan, R., & Saccuzzo, D. P. (1993). *Psychological testing: Principles, applications, and issues* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Larson, G. E., Alderton, D. L., & Kaupp, M. A. (1991). Dynamic administration of a general intelligence test. *Learning and Individual Differences, 3*, 123-134.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*, 107-127.

Raven, J. (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. (1986). Manual for Raven's progressive matrices and vocabulary scales. Research Suppl. No. 3: A compendium of North American normative and validity studies. London: H. K. Lewis. San Antonio, TX: The Psychological Corporation.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 2, pp. 47-103). Hillsdale, NJ: Lawrence Erlbaum.

Spearman, C. (1927). The abilities of man. New York: Macmillan.