# Comparison of Black-White Differences on the WISC-R and the K-ABC: Spearman's Hypothesis

JACK A. NAGLIERI

*Ohio State University*

ARTHUR R. JENSEN

*University of California, Berkeley*

The Kaufman Assessment Battery for Children (K-ABC) and the Wechsler Intelligence Scale for Children-Revised (WISC-R) are compared with respect to the magnitudes of the average white-black differences in standardized scaled scores and in raw scores. The two test batteries were administered to a sample of 172 fourth- and fifth-grade children comprising 86 black-white pairs matched on age, sex, school, and socioeconomic status. The K-ABC and WISC-R are highly correlated, and the general factor, or $g$, of one battery is virtually identical to the $g$ of the other. The high positive correlation between the size of the white-black difference on the various subtests of both batteries and the subtests' loadings on the $g$ factor bears out Spearman's hypothesis that a test's white-black discriminability is a direct function of the test's $g$ loading. The lesser white-black discriminability of the K-ABC relative to the WISC-R is attributable to (1) the smaller $g$ loadings of the K-ABC subtests and (2) the presence of other factors, particularly sequential short-term memory which, to some degree, offsets the white-black difference in $g$.

The Kaufman Assessment Battery for Children (K-ABC) (Kaufman & Kaufman, 1983) owes some of its initial prominence and popularity to its claim that the magnitude of the average black-white difference on the K-ABC IQ is smaller (by about one-half) than the average black-white IQ difference generally reported for most other standardized tests of intelligence, such as the Stanford-Binet and the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974), which is the most widely used individual IQ test in today's schools.

As Jensen (1984) and Naglieri (1986) have previously pointed out, the claim of a lesser black-white difference on the K-ABC was based on evidence that is methodologically far from optimal for establishing this claim. In comparisons of the K-ABC with other tests, such as the WISC-R, both tests were not administered to the same black and white samples. Moreover, comparisons were based on the scaled scores derived from the different standardization samples of the K-ABC and the WISC-R, which renders ambiguous the locus of the tests' dif-

---

ference in black-white discriminability, that is, whether the difference in discriminability is intrinsic to the nature of the tests themselves or is merely an artifact attributable to differences in the range of ability represented in the tests' standardization samples.

By testing the same samples of black and white children on both the K-ABC and the WISC-R, Naglieri (1986) found that, although the white sample obtained higher scaled scores than the black sample on all 11 of the WISC-R subtests and on all 13 of the K-ABC subtests, the sizes of the black-white differences generally were slightly smaller on the K-ABC than on the WISC-R, but not smaller to the degree claimed by the authors of the K-ABC (Kaufman & Kaufman, 1983). This smaller effect, however, is probably due in some part to the fact that Naglieri's black and white samples were closely matched on socioeconomic status. Because the comparisons were made using the standardized scaled scores of each test, the question of the degree of intrinsic difference between the K-ABC and WISC-R with respect to black-white discriminability essentially remained unanswered. Black-white differences in standardized scores on the two tests reflect to some unknown degree possible differences in the tests' standardization samples.

The purpose of the present study is to throw light on this question by basing comparisons on both standardized scores and raw scores. In addition, the study seeks to elucidate the basis for the K-ABC's lesser black-white difference in terms of the hypothesis originally attributable to Charles Spearman (1927), who suggested that tests differ in their black-white discriminability as a positive function of their loadings on the general factor, or $g$, defined as the largest common factor in a collection of diverse cognitive tests. Accordingly, Jensen (1984) has argued that the lesser black-white discriminability of the K-ABC, relative to the WISC-R, is due to the smaller $g$ saturation of the K-ABC, particularly the eight subtests of the K-ABC Mental Processing scale on which the IQ is based. The present study tests both Spearman's hypothesis and Jensen's application of it to understanding the nature of the difference between the K-ABC and the WISC-R with respect to their black-white discriminability.

## METHOD

### Subjects
The sample comprised a total of 172 elementary school pupils in the fourth and fifth grades of three elementary schools in the central district of Columbus, OH. These three schools were selected because they enrolled approximately equal numbers of black and white children, which made it feasible to match the black and white samples on relevant variables. A matched-pairs procedure was used. Each black child ($N = 86$) was matched with a white child on age ($\pm 3$ months), school, sex, and socioeconomic status (SES), resulting in 86 black-white matched pairs (40 males, 46 females). A 5-point scale of SES similar to the one presented in the WISC-R Manual (Wechsler, 1974, p. 18) was adopted. It is

based mainly on the parent's occupational level (the higher of the two levels if both parents were employed). Children whose parents were unemployed or had very low incomes were classed as SES Level 4 if the child was enrolled in the reduced-cost lunch program; those enrolled in the free lunch program were classed as SES Level 5. The mean SES for whites was 3.4 (*SD* = 1.2) and for blacks, 3.4 (*SD* = 1.2). The mean age for whites was 10.7 years (*SD* = 8.2 months; range = 9.3–12.4 years), and for blacks, 10.8 years (*SD* = 8.0 months; range 9.4–12.4 years). The distribution of blacks and whites in the five levels of SES is as follows:

| SES Level | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| White *N* = | 7 | 17 | 16 | 30 | 16 |
| Black *N* = | 7 | 17 | 14 | 31 | 17 |

## Tests

The WISC-R and K-ABC were individually administered in two sessions a week apart by two white examiners, one male, one female. Each examiner tested the same matched black and white children on both the WISC-R and K-ABC. The two tests were given in a completely counterbalanced order (by race), with the matched pairs as the unit of assignment to the two order conditions; the members of any given pair were always given the two tests in the same order.

The WISC-R subtests are familiar enough to the readers of this journal not to need description here. The subtests of the newer and less well-known K-ABC, however, warrant brief descriptions:

## Mental Processing Scale

*Hand movements* (HM) Imitating a series of three hand movements in the same sequence as the examiner performed them.

*Number recall* (NR) Repeating a series of digits in the same sequence as the examiner said them.

*Word order* (WO) Touching a series of pictures in the same sequence as they were named by the examiner, with more difficult items employing a color interference task.

*Gestalt closure* (GC) Naming the object or scene pictured in a partially completed "inkblot" drawing.

*Triangles* (T) Assembling several identical triangles into an abstract pattern that matches a model.

*Matrix analogies* (MA) Selecting the picture or abstract design that best completes a visual analogy.

*Spatial memory* (SM) Recalling the placement of pictures on a page that was exposed briefly.

*Photo series* (PS) Placing photographs of an event in chronological order.

## Achievement Subtests

*Faces & places* (FP) Naming the well-known person, fictional character, or place pictured in a photograph or illustration.

*Arithmetic* (AR) Answering a question that requires knowledge of math concepts or the manipulation of numbers.

*Riddles* (R) Naming the object or concept described by a list of its characteristics.

*Reading/decoding* (RD) Naming letters and reading words.

*Reading/understanding* (RU) Acting out commands given in written sentences.

## RESULTS AND DISCUSSION

### Comparison of Black-White Differences on WISC-R and K-ABC

Table 1 shows the mean raw scores and the standardized scores (based on the national standardization sample) for the black and white samples. For both the raw scores and standardized scores, the mean difference between the black and white samples is expressed in units of the average within-groups standard deviation, labeled $\sigma$ Diff. (see Footnote *a* in Table 1).

Several points in Table 1 should be noted: The raw score black-white differences ($\sigma$ Diff.) are very highly correlated with the standardized score differences, .984 for the WISC-R and .935 for the K-ABC. (The corresponding rank-order correlations are .975 and .960.) The lower correlation for the K-ABC is probably not a statistically significant deviation and seems too slight to warrant any interpretation.

How similar is the pattern of the mean black-white differences on the 11 WISC-R subtests (standardized scores) in the present study to the corresponding pattern of black-white differences found in the national standardization sample of the WISC-R? The rank-order correlation between the two patterns (i.e., the two analogous sets of mean white-black differences on the 11 WISC-R subtests) is +.72. The same question was asked concerning the mean white-black differences on the 13 subtests of the K-ABC. The pattern of the 13 white-black differences (in standard scores) found in the present study has a rank-order correlation of +.81 with the corresponding pattern of white-black differences in the national standardization sample of the K-ABC. Hence, the patterns of white-black differences across the subtests of the WISC-R and the K-ABC in the present samples are not atypical of the corresponding patterns found in the national standardization samples. This fact enhances the generalizability of the present findings, although the present samples are matched on schools and SES, whereas the standardization samples are not.

Although the pattern of differences is much the same for both the raw scores and the standardized scores, the white-black differences differ in absolute level

## TABLE 1
### Summary Statistics on WISC-R and K-ABC in Black ($N = 86$) and White ($N = 86$) Samples

| | Raw Scores | | | | | Standardized Scores | | | | | g Loading | |
| | White | | Black | | | White | | Black | | | | |
| Subtest | Mean | SD | Mean | SD | σ Diff.[a] | Mean | SD | Mean | SD | σ Diff.[a] | White | Black |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **WISC-R** | | | | | | | | | | | | |
| I - Information | 14.17 | 2.50 | 12.94 | 2.57 | .49 | 9.83 | 2.11 | 8.58 | 2.47 | .54 | .591 | .618 |
| S -Similarities | 14.57 | 3.54 | 12.29 | 3.13 | .68 | 10.73 | 2.91 | 8.87 | 2.66 | .67 | .623 | .623 |
| A - Arithmetic | 11.66 | 1.92 | 10.88 | 1.91 | .41 | 9.53 | 2.36 | 8.51 | 2.40 | .43 | .500 | .498 |
| V - Vocabulary | 31.88 | 5.35 | 28.92 | 5.17 | .56 | 10.35 | 2.40 | 9.05 | 2.49 | .53 | .669 | .619 |
| C - Comprehension | 18.30 | 4.32 | 16.39 | 3.81 | .47 | 10.29 | 2.82 | 9.05 | 2.41 | .47 | .520 | .670 |
| DS - Digit Span | 10.52 | 2.74 | 10.26 | 2.60 | .10 | 8.93 | 2.90 | 8.52 | 2.72 | .15 | .423 | .418 |
| PC - Picture Completion | 18.20 | 2.75 | 17.22 | 2.72 | .36 | 10.09 | 2.53 | 9.08 | 2.38 | .41 | .455 | .503 |
| PA - Picture Arrangement | 29.59 | 5.39 | 27.61 | 6.31 | .34 | 11.38 | 2.12 | 10.55 | 2.33 | .38 | .334 | .495 |
| BD - Block Design | 27.43 | 10.61 | 20.16 | 10.02 | .70 | 10.13 | 2.71 | 7.78 | 3.16 | .80 | .627 | .686 |
| OA - Object Assembly | 20.88 | 4.69 | 17.52 | 4.97 | .70 | 9.98 | 3.02 | 8.00 | 2.98 | .66 | .534 | .397 |
| Cd - Coding | 43.12 | 8.27 | 43.29 | 7.36 | -.02 | 10.37 | 2.77 | 10.20 | 2.39 | .07 | .230 | .285 |
| **K-ABC: Mental Processing** | | | | | | | | | | | | |
| HM - Hand Movements | 12.65 | 2.73 | 12.00 | 4.38 | .18 | 9.09 | 2.34 | 8.10 | 2.24 | .43 | .457 | .377 |
| GC - Gestalt Closure | 18.72 | 2.45 | 18.36 | 2.79 | .14 | 10.07 | 2.72 | 9.40 | 2.95 | .24 | .318 | .403 |
| NR - Number Recall | 11.01 | 2.06 | 10.93 | 1.90 | .04 | 9.84 | 2.52 | 9.79 | 2.47 | .02 | .324 | .363 |
| T - Triangles | 14.48 | 2.28 | 13.28 | 2.02 | .56 | 10.56 | 2.69 | 9.03 | 2.32 | .61 | .572 | .521 |
| WO - Word Order | 13.03 | 2.04 | 12.86 | 1.54 | .10 | 9.19 | 2.12 | 8.87 | 1.67 | .17 | .392 | .357 |
| MA - Matrix Analogies | 13.00 | 3.36 | 11.91 | 3.47 | .32 | 9.53 | 2.76 | 9.03 | 2.32 | .47 | .541 | .624 |
| SM - Spatial Memory | 14.87 | 2.47 | 13.81 | 2.30 | .44 | 9.53 | 2.51 | 8.42 | 2.43 | .45 | .349 | .485 |
| PS - Photo Series | 12.80 | 2.12 | 12.38 | 2.40 | .19 | 10.07 | 2.04 | 9.55 | 2.21 | .25 | .411 | .399 |

25

**TABLE 1** (*Continued*)

| Subtest | Raw Scores | | | | | Standardized Scores | | | | | g Loading | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | White | | Black | | | White | | Black | | | White | Black |
| | Mean | SD | Mean | SD | σ Diff.[a] | Mean | SD | Mean | SD | σ Diff.[a] | | |
| **K-ABC: Achievement** | | | | | | | | | | | | |
| FP - Faces and Places | 19.05 | 4.26 | 18.03 | 3.61 | .26 | 99.63 | 12.55 | 96.06 | 12.44 | .29 | .563 | .531 |
| AR - Arithmetic | 29.65 | 3.99 | 27.48 | 2.68 | .64 | 99.27 | 12.17 | 90.53 | 9.40 | .80 | .621 | .514 |
| R - Riddles | 22.88 | 3.71 | 20.58 | 3.43 | .64 | 99.80 | 11.08 | 91.95 | 11.42 | .70 | .649 | .694 |
| RD - Reading/Decoding | 30.06 | 3.53 | 29.30 | 2.72 | .24 | 100.80 | 12.69 | 95.50 | 11.05 | .42 | .614 | .516 |
| RU - Reading/Understanding | 15.85 | 2.49 | 14.94 | 2.45 | .37 | 96.66 | 9.02 | 92.74 | 8.45 | .45 | .571 | .605 |

| Scales | Raw Scores | | | | | Standardized Scores | | | | | Mean g | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | White | | Black | | | White | | Black | | | White | Black |
| | Mean | SD | Mean | SD | σ Diff.[a] | Mean | SD | Mean | SD | σ Diff.[a] | | |
| WISC-R: Verbal | 101.12 | 15.20 | 91.69 | 14.18 | .64 | 100.53 | 12.13 | 92.41 | 12.56 | .66 | .554 | .574 |
| WISC-R: Performance | 139.22 | 21.22 | 125.80 | 21.71 | .57 | 102.28 | 12.46 | 93.67 | 12.62 | .69 | .436 | .473 |
| WISC-R: Full Scale | 240.34 | 31.60 | 217.49 | 31.08 | .73 | 101.38 | 11.65 | 92.30 | 11.84 | .77 | .500 | .528 |
| K-ABC: Mental Processing | 110.57 | 11.62 | 105.53 | 12.30 | .42 | 97.56 | 10.68 | 91.53 | 10.15 | .58 | .421 | .441 |
| K-ABC: Achievement | 117.49 | 14.07 | 110.34 | 11.20 | .56 | 98.87 | 10.79 | 92.03 | 9.60 | .67 | .604 | .572 |
| K-ABC: Total | 228.06 | 23.23 | 215.87 | 20.19 | .56 | —[b] | | — | | — | .491 | .491 |

[a] σ Diff. is the mean white-black difference divided by $\bar{\sigma}$, where $\bar{\sigma} = \sqrt{(s_W^2 + s_B^2)/2}$. For 170 $df$ a σ Diff. greater than 0.30 is significant beyond the .05 level (2-tailed test).

[b] The K-ABC Manual does not give standardized scores for the composite of all the subtests.

for the two types of scores. The absolute difference between the raw score and standardized score differences is greater for the K-ABC than for the WISC-R. The mean (across subtests) white-black differences (in σ units) for raw and standardized scores are as follows:

|  | Raw | Standardized |
|---|---|---|
| WISC-R | .435σ | .464σ |
| K-ABC Mental Processing | .246σ | .328σ |
| K-ABC Achievement | .430σ | .531σ |
| K-ABC Total | .317σ | .406σ |

This result contradicts Jensen's (1984) conjecture that a part of the lesser white-black difference on the K-ABC than on the WISC-R might be attributable to differences in the variance of ability in the different standardization populations of the WISC-R and K-ABC. In fact, in the present samples, the white-black difference on K-ABC standardized scores in *larger* than on the raw scores, and the K-ABC differs more from the WISC-R on the raw score white-black differences than on the standardized score differences. The standardization of the K-ABC scores, therefore, appears not to be one of the mechanisms responsible for the generally smaller white-black difference on the K-ABC relative to the white-black difference on the WISC-R.

The same effects are seen in the WISC-R and K-ABC scale composites shown at the bottom of Table 1. The K-ABC Mental Processing composite has a white-black difference only 57% (i.e., 100 × .42σ/.73σ) as large as the Full Scale WISC-R for raw scores and 75% (i.e., 100 × .58σ/.77σ) as large for standardized scores. In terms of the tests' actual IQ scales, the white-black difference on the Full Scale WISC-R is 9.1 IQ points; on the K-ABC Mental Processing Scale it is 6.0 IQ points, or two-thirds the size of the difference on the WISC-R. The K-ABC Achievement Scale shows a white-black difference equivalent to 6.8 IQ points, although the K-ABC does not use the achievement subtests in computing the IQ, which is based entirely on the mental processing subtests. But, the main point to be noted here is that the K-ABC yields smaller white-black differences, expressed in σ units, than the Full Scale WISC-R, for all of the K-ABC scales, whether compared in terms of raw scores or standardized scores. However, the raw-score comparisons are the more essential for our purpose, because they are wholly unaffected by any possible differences in the standardization samples of the WISC-R and the K-ABC.

**Correlations between WISC-R and K-ABC**
The Pearson correlation coefficients between all of the WISC-R and K-ABC subtests are shown separately for the black and the white samples in Table 2. To determine whether the correlation matrices of the two samples differ significantly, the Jennrich (1970) chi-square test for the equality of two correlation

# TABLE 2

**Correlations (Decimals Omitted) Between Subtests of the WISC-R and K-ABC in Black (Above Diagonal) and White (Below Diagonal) Samples**

| Subtest | WISC-R Verbal | | | | | | WISC-R Performance | | | | | K-ABC Mental Processing | | | | | | | | K-ABC Achievement | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | I | S | A | V | C | DS | PC | PA | B | OA | Cd | HM | GC | NR | T | WO | MA | SM | PS | FP | Ar | R | R/D | R/U |
| Information | | 71 | 48 | 70 | 68 | 11 | 39 | 36 | 45 | 20 | 11 | 15 | 25 | 16 | 27 | 24 | 44 | 26 | 19 | 66 | 44 | 73 | 55 | 59 |
| Similarities | 60 | | 46 | 65 | 67 | 19 | 43 | 41 | 56 | 25 | 04 | 18 | 33 | 17 | 40 | 20 | 48 | 26 | 18 | 53 | 45 | 64 | 43 | 56 |
| Arithmetic | 37 | 37 | | 44 | 53 | 15 | 23 | 30 | 44 | 06 | 21 | 27 | 18 | 19 | 27 | 25 | 35 | 18 | 20 | 25 | 59 | 43 | 35 | 41 |
| Vocabulary | 71 | 69 | 38 | | 76 | 25 | 26 | 33 | 43 | 16 | 13 | 11 | 26 | 27 | 31 | 10 | 43 | 22 | 14 | 69 | 47 | 69 | 60 | 67 |
| Comprehension | 59 | 51 | 25 | 63 | | 23 | 39 | 39 | 49 | 19 | 18 | 21 | 27 | 23 | 33 | 28 | 44 | 28 | 26 | 55 | 57 | 67 | 48 | 61 |
| Digit Span | 20 | 23 | 52 | 11 | 10 | | 10 | 17 | 22 | 12 | 24 | 38 | 18 | 52 | 19 | 29 | 33 | 31 | 12 | 18 | 14 | 24 | 30 | 15 |
| Picture Completion | 29 | 38 | 09 | 28 | 34 | 14 | | 39 | 58 | 44 | 09 | 16 | 41 | 14 | 51 | 25 | 45 | 33 | 38 | 35 | 27 | 40 | 17 | 27 |
| Picture Arrangement | 24 | 17 | 23 | 28 | 11 | 14 | 18 | | 44 | 37 | 21 | 23 | 36 | 06 | 39 | 23 | 26 | 35 | 44 | 34 | 26 | 58 | 10 | 27 |
| Block Design | 38 | 41 | 34 | 56 | 23 | 25 | 43 | 40 | | 51 | 25 | 29 | 42 | 23 | 66 | 18 | 59 | 44 | 45 | 38 | 51 | 55 | 34 | 49 |
| Object Assembly | 29 | 37 | 24 | 45 | 24 | 23 | 44 | 29 | 55 | | 27 | 15 | 35 | 01 | 53 | 08 | 38 | 48 | 35 | 27 | 06 | 28 | 11 | 21 |

28

| Coding | 03 | 15 | 10 | 05 | 05 | 14 | 15 | 22 | 16 | 13 | 26 | 10 | 08 | 12 | 12 | 22 | 30 | 39 | 09 | 07 | 17 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hand Movements | 19 | 36 | 45 | 26 | 31 | 40 | 16 | 05 | 21 | 25 | | 13 | 19 | 19 | 26 | 26 | 22 | 28 | 13 | 28 | 20 | 16 | 25 |
| Gestalt Closure | 39 | 31 | 00 | 28 | 30 | 11 | 37 | 13 | 36 | 26 | 00 | | 06 | 50 | 08 | 34 | 27 | 21 | 29 | 11 | 37 | 22 | 33 |
| Number Recall | 12 | 12 | 29 | 13 | 11 | 63 | 13 | 14 | 30 | 18 | 35 | −08 | | 22 | 40 | 24 | 18 | 09 | 09 | 18 | 26 | 23 | 24 |
| Triangles | 31 | 29 | 44 | 37 | 21 | 26 | 40 | 37 | 674 | 53 | 31 | 07 | 25 | | 10 | 41 | 48 | 28 | 26 | 30 | 39 | 12 | 23 |
| Word Order | 11 | 18 | 39 | 12 | 17 | 64 | 19 | 04 | 29 | 21 | 36 | −06 | 59 | 31 | | 32 | 30 | 11 | 09 | 23 | 18 | 12 | 13 |
| Matrix Analogies | 23 | 36 | 39 | 41 | 38 | 26 | 36 | 32 | 48 | 40 | 39 | 17 | 21 | 44 | 25 | | 41 | 35 | 35 | 41 | 47 | 52 | 46 |
| Spatial Memory | 02 | 14 | 09 | 07 | 09 | 14 | 29 | 20 | 28 | 32 | 30 | 17 | 09 | 39 | 25 | 29 | | 38 | 24 | 12 | 30 | 24 | 21 |
| Photo Series | 31 | 40 | 32 | 30 | 31 | 09 | 17 | 10 | 29 | 32 | 33 | 32 | −02 | 31 | 10 | 14 | 29 | | 12 | 21 | 31 | 10 | 19 |
| Faces and Places | 72 | 68 | 20 | 75 | 54 | 09 | 34 | 20 | 40 | 35 | 24 | 35 | 08 | 24 | 07 | 26 | −03 | 32 | | 29 | 69 | 59 | 67 |
| Arithmetic | 49 | 44 | 57 | 54 | 49 | 43 | 15 | 24 | 44 | 24 | 46 | 13 | 34 | 35 | 35 | 42 | 29 | 33 | 41 | | 48 | 39 | 45 |
| Riddles | 66 | 64 | 33 | 73 | 49 | 15 | 38 | 24 | 46 | 45 | 34 | 39 | 10 | 35 | 14 | 33 | 12 | 42 | 70 | 48 | | 52 | 63 |
| Reading/Decoding | 53 | 50 | 41 | 63 | 48 | 41 | 35 | 14 | 42 | 30 | 40 | 12 | 35 | 30 | 29 | 51 | 10 | 17 | 53 | 65 | 55 | | 73 |
| Reading Understanding | 55 | 51 | 39 | 59 | 46 | 27 | 23 | 18 | 33 | 29 | 34 | 09 | 16 | 36 | 25 | 37 | 14 | 30 | 47 | 53 | 60 | 68 | |

matrices was applied. The resulting $\chi^2$, with 276 *df*, is 1817.95, indicating a difference that is significant beyond the .001 level. This means, too, that the factor structure of the combined WISC-R and K-ABC batteries could be significantly different between the black and white samples, although the few largest and significant factors possibly might not differ appreciably.

The general factor common to all of the subtests of both batteries, however, is highly similar—in fact, virtually identical—in both samples. As it is the general factor that constitutes most of the variance in the composite scale scores of each battery, it is not surprising that the WISC-R and K-ABC scales show high and similar correlations in both samples, as shown in Table 3.

The high correlations between the K-ABC Mental Processing and Achievement scales with the WISC-R Full Scale IQ presented in Table 3 demonstrate that these scales are both more highly correlated with the WISC-R Full Scale than they are with each other (K-ABC Mental Processing × Achievement correlations are .50, .58, and .58 for the black, white, and total samples, respectively). This finding is further supported by Naglieri (1985a), who found that the WISC-R Full Scale IQ is correlated significantly higher with the K-ABC Achievement scale than with the K-ABC Mental Processing scale. Additionally, Naglieri (1984), Naglieri and Anderson (1985), and Naglieri (1985a, 1985b) found the K-ABC Achievement and the WISC-R Verbal to be the most highly correlated of the separate scales of the two tests. Because these scales essentially measure verbal knowledge acquired through educational and other means, this result is expected.

The strong correlations between the WISC-R Performance and K-ABC Simultaneous scales evident in Table 3 reflect the nonverbal spatial nature of the tasks on each of the scales. This strong Simultaneous/Performance relationship

TABLE 3
Correlations between WISC–R and K–ABC Scales for Black $N = 86$), White ($N = 86$), and Total ($N = 172$) Samples

| WISC–R Scales | | K–ABC Scales | | | |
|---|---|---|---|---|---|
| | | Simultaneous | Sequential | Mental Processing | Achievement |
| Verbal | Black | .500 | .325 | .523 | .826 |
| | White | .552 | .350 | .562 | .848 |
| | Total | .566 | .359 | .582 | .852 |
| Performance | Black | .793 | .326 | .750 | .496 |
| | White | .758 | .327 | .693 | .466 |
| | Total | .797 | .349 | .745 | .533 |
| Full scale | Black | .733 | .374 | .726 | .770 |
| | White | .747 | .400 | .723 | .771 |
| | Total | .765 | .406 | .749 | .795 |

also influences the correlations between the Mental Processing and Performance scales because the K-ABC Mental Processing scale is simultaneously weighted. That is, the Mental Processing scale comprises five Simultaneous and three Sequential subtests, one of which—Hand Movements—evidenced strong simultaneous loadings (Kaufman & Kamphaus, 1985), that contribute equally to the scale. These results indicate that the WISC-R and K-ABC scales are consistent in the measurement of verbal/achievement and nonverbal/spatial intellectual skills.

The above correlations may be compared with the correlations between other IQ tests. In a quite comprehensive review of the correlations between many different intelligence tests, it was found that the overall average correlation is $+.67$ (Jensen, 1980, pp. 314–316). Most such correlations fall in the range $+.50$ to $+.80$. In 47 studies reporting a correlation between the WISC IQ and the Stanford-Binet IQ, the correlations range from $+.43$ to $+.94$, with a median correlation of $+.80$. Hence, the K-ABC is not atypical in its correlations with the WISC-R. If the K-ABC based its IQ scale on all 13 of the subtests, rather than on just the 8 Mental Processing subtests, it would correlate as highly with the WISC-R Full Scale IQ as does the Stanford-Binet IQ. The K-ABC Mental Processing and Achievement scores together predict WISC-R Full Scale IQ with the following multiple correlations: black $= .87$, white $= .84$, and combined $= .87$. Corrected for attenuation, assuming a reliability coefficient of .90 for each of the three scales, these $R$s would become .91, .95, and .95, respectively. Hence, the WISC-R and K-ABC scales obviously have a great deal of their true-score variance in common, undoubtedly because the composite score of each battery measures predominantly the same general factor.

The variance of group differences, as of individual differences, in the total scores of a test battery reflects the sum of the separate subtest variances plus twice the sum of all the covariances among the subtests. Consequently, a test battery that has generally lower intercorrelations among its subtests will be somewhat less reliably discriminating between individuals and between groups than a battery with higher subtest intercorrelations. The average intercorrelations among subtests of the WISC-R and of the K-ABC scales in the black and white samples are:

|  | Black | White |
| --- | --- | --- |
| WISC-R | $+.342$ | $+.298$ |
| K-ABC Mental Processing | $+.259$ | $+.233$ |
| K-ABC Achievement | $+.544$ | $+.560$ |
| K-ABC All Subtests | $+.334$ | $+.319$ |

## Factor Analysis
A principal factor analysis was performed on the intercorrelations among all 24 of the WISC-R and K-ABC subtests, separately in the black and white groups

(Table 2) and in the combined groups. So that the white-black differences would not inflate the correlations for the combined groups, the point-biserial correlations of the dichotomized race variable with each of the 24 subtests were used to partial out race from all of the subtest intercorrelations. (When the two combined groups are of equal size, as is the case here, the correlation, $r_{xy'}$ between any two variables, X and Y, with the group difference, Z, partialled out, i.e., $r_{xy \cdot z'}$ is equal to the mean of the separate zero-order correlations, $r_{xy'}$ in each of the groups.) It is essential for the subsequent test of Spearman's hypothesis concerning the factorial composition of the white-black difference on psychometric tests that the factor analysis of the tests not be the least contaminated by the group difference itself.

Three criteria were appealed to for determining the number of factors: the Kaiser-Guttman rule of extracting only factors with eigenvalues greater than 1, Cattell's scree test, and a maximum likelihood factor analysis, using the LISREL program (Jöreskog & Sörbom, 1981), to test the goodness of fit of the three-factor model (see Gorsuch, 1983, Chapter 8). According to each of the three criteria, for all three correlation matrices (white, black, and combined), the correct number of factors is three. Two factors are clearly inadequate and, when four factors are extracted, the fourth factor, after orthogonal rotation, consists of merely a doublet, with quite small loadings on only the two arithmetic tests— one in the WISC-R and one in the K-ABC. Also, the maximum likelihood test yields a better goodness-of-fit index for the three-factors than for the four-factors solution.

The first three principal factors, then, were used to obtain a Schmid–Leiman (1957) orthogonalized hierarchical factor analysis in the separate and in the combined groups.[1] The results are shown in Table 4.

---

[1] The orthogonalized hierarchical factor analysis was performed by means of an SAS computer program based on the routine developed by Wherry (1959), which makes use of multiple group factor analysis and obviates the need for oblique rotation of the primary factors (derived in the present case by principal factor analysis). The simple structure criterion is approximated by removing the overlapping variance between the clusters identified by the principal factor analysis and expressing it as a higher order factor, in the present case, the second-order $g$ factor. The final result—an orthogonalized hierarchical factor structure—is exactly the same factor matrix that is yielded by the method described by Schmid and Leiman (1957).

Although correlations between the oblique primary factors are not required for obtaining the final orthogonalized hierarchical solution by Wherry's method, Wherry provides the routine for calculating them as an optional by-product of his method. In the present study, the correlations between the Verbal (V), Spatial (S), and Memory Span (MS) oblique primary factors in the white, black, and combined samples are:

| Correlation | White | Black | Combined |
|---|---|---|---|
| V–S | .62 | .61 | .61 |
| V–MS | .48 | .57 | .52 |
| S–MS | .48 | .57 | .51 |

The three primary factors, Verbal (V), Spatial (S) and Memory Span (MS), are especially clear after residualization by the Schmid-Leiman procedure. Moreover, the subtests with salient loadings (in italics) on each factor are highly similar in the white, black, and combined groups, indicating essentially the same factor structure in both racial groups. The second-order or general factor, g, on which our subsequent analyses mainly focus, shows a coefficient of congruence[2] of +.95 between the black and white groups, indicating that this g factor is essentially the same in both groups. Also, a second-order general factor was extracted from the combined data (with the group difference partialled out) by means of maximum likelihood factor analysis. It shows a congruence coefficient of +.99 with the Schmid-Leiman g factor extracted from the same combined data. As can be seen in the bottom line of Table 4, in every analysis the second-order g factor accounts for more than twice as much of the common factor variance as the three first-order factors combined.

### Concurrent Validity of the g Factor

Because the WISC-R and K-ABC subtests have not heretofore been factor analyzed together, we cannot compare the g of the present analysis with that of any other analysis. Yet, one might wonder if the present g bears much resemblance to the g of the WISC-R and K-ABC found in other samples. It turns out that factor analyzing the two batteries together creates very little distortion of the pattern of g factor loadings in either battery as compared to the g loadings obtained when the two batteries are each factor analyzed separately. In the present combined samples, the g factor loadings obtained from factor analyzing the WISC-R and the K-ABC separately were compared with the corresponding g loadings obtained from factor analyzing the two test batteries together. The coefficients of congruence between the two sets of g loadings thus obtained are +.998 for both the WISC-R and the K-ABC, indicating virtual identity of the g factor whether the WISC-R and K-ABC batteries are factor analyzed separately or combined.

Even when factor analysis is performed on the two batteries separately, and in entirely *different samples,* the resulting g factors are highly congruent with the g of the combined batteries in the present samples. It is striking evidence of the ubiquity and robustness of g. For example, the g loadings of the WISC-R

---

[2]The coefficient of congruence, $r_c$, is an index of factor similarity on a scale of 0 to ± 1. Unlike the Pearson r, which, being based on standardized variates, reflects only the degree of similarity between the profiles (of factor loadings) per se, the congruence coefficient also reflects differences in the absolute values of the factor loadings. A value of $r_c$ above +.90 is the usual criterion for concluding identity of factors, although some experts set a more stringent criterion at +.95. The congruence coefficient is computed as follows:

$$r_c = \frac{\Sigma ab}{\sqrt{\Sigma a^2 \ \Sigma b^2}}$$

where a and b are the homologous factor loadings obtained on a given factor in groups A and B.

TABLE 4
Schmid-Leiman Orthogonalized Hierarchical Factor Analysis of WISC-R and K-ABC
Subscales Showing Loadings (Decimals Omitted) of General (g), Verbal (V), and Memory
Span (MS) Factors in the White, Black, and Combined Samples

| Subtest | White (N = 86) | | | | Black (N = 86) | | | | Combined (N = 172) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | g | V | S | MS | g | V | S | MS | g | V | S | MS |
| I | 591 | 550 | −031 | −097 | 618 | 554 | −009 | −052 | 603 | 547 | −014 | −081 |
| S | 623 | 421 | 072 | −015 | 623 | 437 | 092 | −030 | 623 | 426 | 082 | −022 |
| A | 500 | 135 | 018 | 390 | 498 | 292 | 005 | 121 | 504 | 205 | 008 | 266 |
| V | 669 | 577 | 034 | −148 | 619 | 609 | −076 | −037 | 640 | 597 | −022 | −099 |
| C | 520 | 435 | 000 | −050 | 670 | 473 | 001 | 077 | 583 | 433 | 010 | −005 |
| DS | 423 | −071 | −028 | 710 | 418 | −101 | −007 | 514 | 421 | −082 | −017 | 619 |
| PC | 455 | 069 | 337 | −089 | 503 | 049 | 421 | −074 | 477 | 059 | 382 | −085 |
| PA | 334 | −001 | 281 | −033 | 495 | 086 | 316 | −002 | 416 | 043 | 309 | −027 |
| BD | 627 | 053 | 418 | 027 | 686 | 115 | 451 | −013 | 656 | 079 | 441 | 006 |
| OA | 534 | 018 | 395 | 000 | 397 | −092 | 591 | −205 | 472 | −043 | 481 | −083 |
| Cd | 230 | −211 | 362 | 043 | 285 | −113 | 190 | 177 | 245 | −153 | 281 | 101 |
| HM | 457 | 098 | 045 | 353 | 377 | −081 | 097 | 332 | 415 | 017 | 077 | 338 |
| GC | 318 | 149 | 254 | −259 | 403 | 062 | 343 | −092 | 356 | 107 | 307 | −183 |
| NR | 324 | −059 | −061 | 619 | 363 | −029 | −112 | 500 | 342 | −041 | −083 | 564 |
| T | 572 | −079 | 440 | 137 | 521 | −043 | 500 | −041 | 549 | −065 | 455 | 074 |
| WO | 392 | −102 | 026 | 631 | 357 | −081 | 017 | 406 | 382 | −098 | 014 | 553 |
| MA | 541 | 079 | 259 | 136 | 624 | 149 | 240 | 133 | 578 | 115 | 254 | 131 |
| SM | 349 | −225 | 465 | 050 | 485 | −109 | 401 | 114 | 410 | −160 | 438 | 075 |
| PS | 411 | 149 | 179 | −013 | 399 | −104 | 422 | 004 | 403 | 018 | 310 | −008 |
| FP | 563 | 591 | −039 | −187 | 531 | 551 | 034 | −180 | 549 | 569 | −006 | −185 |
| AR | 621 | 298 | −003 | 313 | 514 | 267 | −035 | 212 | 573 | 283 | −026 | 281 |
| R | 649 | 481 | 090 | −109 | 694 | 465 | 129 | −038 | 670 | 472 | 116 | −082 |
| R/D | 614 | 445 | −094 | 213 | 516 | 468 | −110 | 067 | 573 | 453 | −110 | 157 |
| R/U | 571 | 427 | −030 | 079 | 605 | 529 | −025 | −018 | 592 | 470 | −029 | 037 |
| % Var. | 26.0 | 9.4 | 5.3 | 8.0 | 27.1 | 9.9 | 7.1 | 4.2 | 26.4 | 9.4 | 6.0 | 6.0 |

subtests in the WISC-R national standardization sample (Jensen & Reynolds, 1982), obtained from a Schmid-Leiman hierarchical factor analysis of just the WISC-R subtests, show a congruence coefficient of +.99 with the present $g$ factor loadings of the WISC-R subtests obtained from the present Schmid-Leiman hierarchical factor analysis of the combined WISC-R and K-ABC batteries. The $g$ loadings of the K-ABC subtests in the K-ABC national standardization sample, obtained from a hierarchical factor analysis of just the K-ABC for 10-year-olds (Keith & Dunbar, 1984, Table 2, p. 373), show a congruence coefficient of +.99 with the present $g$ factor loadings of the K-ABC obtained from the combined WISC-R and K-ABC batteries.

Hence, the $g$ factor of the WISC-R and K-ABC in the present samples is

virtually identical to that of the national standardization samples. The fact that combining the two batteries does not distort the pattern of factor loadings within each battery is further evidence that the general factor in both batteries represents essentially one and the same $g$, despite the differences in appearaces between the WISC-R and K-ABC batteries. To determine the similarity of the $g$ factor obtained in separate factor analyses of the WISC-R and K-ABC batteries in the present sample (total $N = 172$), $g$ factor scores for the WISC-R and for the K-ABC were computed on every subject. The Pearson correlation between the $g$ factor scores of the WISC-R and the $g$ factor scores of the K-ABC is $+.88$. As the reliability of each test is probably not higher than .90, the correlation of $+.88$ corrected for attenuation (i.e., $+.88/.90$) yields a true correlation of about $+.98$. Hence, the WISC-R and K-ABC have virtually the same $g$ factor.

## Spearman's Hypothesis

The white-black difference is far from being a constant amount on various kinds of tests. This is clearly evident by inspection of the two $\sigma$ Diff. columns in Table 1, which show the white-black differences vary widely on different subtests. In the 24 subtests of the combined WISC-R and K-ABC batteries, the white-black differences in the present samples range from near zero ($-.02\sigma$) to $+0.70\sigma$. Spearman (1927, p. 379) hypothesized that the observed variation in the size of the black-white difference across different tests is primarily a function of the tests' loadings on the $g$ factor, the more $g$-loaded tests showing the larger differences. Jensen (1985a) has made a detailed examination of 11 large data sets suitable for testing Spearman's hypothesis and found the hypothesis consistently borne out. The overall correlation between $g$ loadings and black-white differences (in standard score units) for 121 tests in 11 studies was $+.59$. Humphreys (1985a), however, has claimed only very weak support (a correlation of only $+.19$ between the $g$ loadings and black-white differences on 74 tests in the Project TALENT battery) for Spearman's hypothesis in a comparison of a representative sample of black high school students with a specially restricted sample of white students of low SES (the lowest 15 to 20% in SES of the total representative sample of white students). The specific features of Humphreys's data from Project TALENT that would account for its unusually weak conformity to Spearman's hypothesis have been pointed out elsewhere (Jensen, 1985b). Humphreys (1985b) later calculated the correlation of tests' $g$ loadings with black-white differences for the total representative samples of blacks and whites, without any restriction on SES, and it turned out to be $+.59$, exactly the same overall correlation found in the 11 studies analyzed by Jensen (1985a).

The question remains open, however, as to how strongly Spearman's hypothesis is borne out, not when one racial group is intentionally restricted in SES and the other is not (which would be an inappropriate test of the hypothesis), but when both groups are *matched* on SES throughout the full range of SES. The present data, which were originally obtained for a study by Naglieri (1986),

permit examination of Spearman's hypothesis for black and white groups matched on SES. They also permit a test of the hypothesis suggested by Jensen (1984) that the K-ABC, particularly the Mental Processing scale, yields a smaller black-white difference than the WISC-R (and many other standardized IQ tests) because the K-ABC subtests are generally less $g$ loaded than those of the WISC-R.

A primary test of Spearman's hypothesis is the correlation between the various tests' $g$ factor loadings and the mean black-white difference (in $\sigma$ units) on the tests. All subsequent analyses are based on the raw scores, so that the black-white differences on the WISC-R and K-ABC will not reflect possible differences in their standardization samples.

The Pearson correlation between $g$ loadings and $\sigma$ differences is $+.78$. Because there is no suitable significance test for the Pearson $r$ in this case,[3] we also report the rank-order correlation, which is $+.75$, $p < .01$, one-tailed test. The scatter diagram for this correlation is shown in Figure 1, indicating the bivariate data points for the various WISC-R and K-ABC subtests.

The correlations between the white-black $\sigma$ differences and the $g$ loadings obtained in the separate groups are $+.71$ for the black and $+.75$ for the white. Over all of the 24 subtests, Spearman's hypothesis appears to be strongly sustantiated.

It is possible, however, that the correlation between $g$ loadings and $\sigma$ differences could reflect merely the differential reliability coefficients of the various subtests. Both $g$ loadings and $\sigma$ differences are attenuated by measurement error, and the degree of attenuation would be correlated between the two variables. To determine if this has substantially affected our test of Spearman's hypothesis, we have controlled for differences in subtest reliabilities in two ways. The most reliable subtest reliabilities available are those computed on the national standardization samples, obtained from the WISC-R and K-ABC test manuals. These were used in the following analyses. First, we partialled the subtest reliabilities out of the correlation between subtest $g$ loadings and $\sigma$ differences. The partial $r$

---

[3]Because the $g$ loadings derived from a particular battery of tests are not statistically independent of one another and do not qualify as a random sample from a population with an assumed normal distribution, and because the same is true of the standardized mean black-white differences on the tests, the Pearson product-moment coefficient of correlation ($r$) between $g$ loadings and mean differences, although it is the most precise index of the degree of linear relationship between the two sets of variables, cannot, in a strict sense, be tested for statistical significance. Therefore, significance tests are not here applied to the Pearson $r$ when used as an index of relationship between $g$ loadings and mean black-white differences. However, in addition to the Pearson $r$, the corresponding Spearman rank-order correlation, $\rho$, is also reported, because its level of significance does not rest on any assumptions about the distributional characteristics of the two variates. As a nonparmetric, or distribution-free, test of independence or index of relationship, the rank correlation's level of significance is simply the proportion $p$ of all possible $n!$ permutations of the $n$-ranked pairs of variables for which the absolute value of $\rho$ is equal to or greater than the obtained $\rho$.
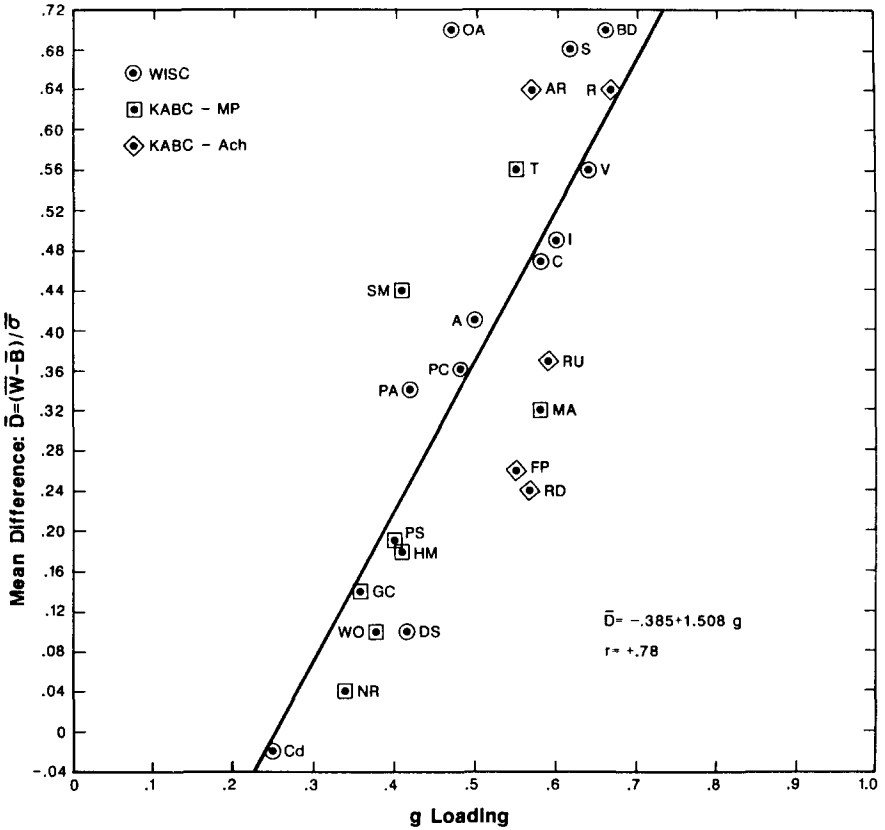
**Fig. 1.** Mean black–white difference (expressed in standard deviation units) on WISC–R and K–ABC subtests as a function of the subtests' loadings on *g*. WISC–R subtests: I—Information, S—Similarities, A—Arithmetic, V—Vocabulary, C—Comprehension, DS—Digit Span, PC—Picture Completion, PA—Picture Arrangement, BD—Block Design, OA—Object Assembly, Cd—Coding. K–ABC Mental Processing (MP) subtests: HM—Hand Movements, GC—Gestalt Closure, NR—Number Recall, T—Triangles, WO—Word Order, MA—Matrix Analogies, SM—Spatial Memory, PS—Photo Series; K–ABC Achievement (Ach) subtests: FP—Faces and Places, Ar—Arithmetic, R—Riddles, RD—Reading/Decoding, RU—Reading/Understanding.

is +.85, which is larger than the zero-order *r* of +.78. Second, we corrected both the *g* loadings and σ differences for attenuation (by dividing each by the square root of the subtest's reliability coefficient). The correlation between the disattenuated *g* loadings and the disattenuated σ differences is +.80. Hence, variation in subtest reliabilities is not in the least responsible for the large positive correlation between subtests' *g* loadings and the mean black-white differences, as predicted by Spearman's hypothesis. Because the disattenuated data yield

scarcely different results from those of the raw data, we will henceforth use only the raw data, unless noted otherwise.

Another way of looking at Spearman's hypothesis, originally suggested by Gordon (1985), is based on the point-biserial correlation between each subtest score and the black-white dichotomy (quantitized as black = 0, white = 1). These point-biserial correlations for the 24 subtests are identical to the subtests' loadings on a black-white factor that would be obtained by including the black-white dichotomy as a variable in a factor analysis of the 24 subtests and rotating the factor axes in such a way as to obtain a maximal loading of 1.00 on the black-white variable. The various subtests' loadings on this "black-white" factor would be the same as the point biserial correlation of each subtest with the dichotomous black-white variable. This black-white factor, then, can be compared with the $g$ factor (as represented by the first principal component) of the 24 subtests by the usual index of factor similarity, the congruence coefficient. This turns out to be +.965, which indicates that the $g$ factor and the black-white factor (as reflected in these 24 subtests) are so similar as to be practically the same factor. The reason for our using the first principal component as the $g$ factor in this context is now explained. Following Gorsuch (1983, p. 285), Gordon (1985) has suggested that one interpretation of the congruence coefficient is that it is equivalent to the value of the Pearson correlation that would obtain between the sets of factor scores derived from each of the two factors that are being compared for similarity. This interpretation of the congruence coefficient appears not to be exactly true. At least, it does not hold for the present data, for which all the required calculations were double-checked. Factor scores[4] on all 172 subjects were calculated from the black-white factor loadings (i.e., the point-biserial correlations between the black-white dichotomy, quantitized as 0 and 1, respectively) on the 24 subtests, and factor scores on the 172 subjects were also calculated from the loadings of the 24 subtests on their first principal component. As previously stated, the congruence coefficient between these two factors is +.965. But, the Pearson correlation between the two sets of factor scores is even higher: +.996. Thus, although Gorsuch's interpretation of the congruence coefficient as an indicator of the correlation between factor scores does not seem to hold, Gordon's argument that the black-white factor is essentially the same as the $g$ factor, as represented by the first principal component, is strongly supported by

---

[4]The black–white factor scores were obtained in essentially the same manner as the first principal component factor scores. That is, a subject's factor score is a weighted average of the subtest scores. The age-standardized scores of all 172 subjects on each of the 24 WISC–R and K–ABC subtests were converted to $z$ scores (mean = 0, $\sigma$ = 1) based on the mean and $\sigma$ of the age-standardized scores in the total sample. A subject's factor score (on the black-white factor) is the weighted mean of the 24 subtest scores, with each of the 24 subtest $z$ scores weighted by the point-biserial correlation between the age-standardized subtest scores and the black–white dichotomy, quantitized as 0 and 1, respectively.

the correlation of $+.996$ between the factor scores based on these two factors. (The congruence coefficient between the first principal component and the Schmid-Leiman second-order $g$ factor is $+.988$.)

The WISC-R and K-ABC can be compared with respect to Spearman's hypothesis. A most conspicuous feature of Figure 1 is that five of the eight Mental Processing tests on which the K-ABC IQ is based cluster in the lower left quadrant, that is, they have relatively low $g$ loadings and relatively small black-white differences. Only the Coding and Digit Span tests of the WISC-R fall into this quadrant. All of the Mental Processing subtests in this quadrant except Gestalt Closure and Photo Series, as well as the two WISC-R subtests, are mainly tests of short-term sequential memory, or what has been termed Level I ability, on which white-black differences have generally been found to be minimal or even reversed when $g$ is partialled out (Vernon, 1981). All of the Level I tests, of course, have some $g$ loading, which accounts for their conformity to Spearman's hypothesis. (The correlation between the $g$ loadings and the mean white-black differences for just these five Level I type tests in the lower left quadrant of Figure 1 is $+.90$.) Hence, the comparatively low $g$ loadings of five of the eight Mental Processing subtests are implicated in the smaller black-white IQ difference on the K-ABC, as compared with the WISC-R IQ. Three of the Mental Processing subtests do not fall into this cluster, and they are all tests that involve some spatial ability factor in addition to $g$, particularly Spatial Memory and Triangles, both of which fall *above* the regression line, that is, the black-white difference is larger than would be predicted from these tests' $g$ loadings. Evidently, another factor (spatial ability) increases the white-black differences on these tests. The same effect is seen for the two most spatially loaded subtests of the WISC-R—Object Assembly and Block Design.

The mean $g$ loadings ($\bar{g}$) and mean black-white differences in $\sigma$ units ($\bar{D}$), along with the conformance ($r$) of the subtests to Spearman's hypothesis for the subtests of the WISC-R and K-ABC scales separately are as follows:

|  | $\bar{g}$ | $\bar{D}$ | $\bar{r}$ |
|---|---|---|---|
| WISC–R | .51 | .43 | .83 |
| K–ABC Mental Processing | .43 | .25 | .74 |
| K–ABC Achievement | .59 | .43 | .62 |
| K–ABC All Subtests | .49 | .32 | .74 |

The K-ABC Achievement tests have nearly the same average $g$ loading as the WISC-R and show the same overall white-black difference as the WISC-R subtests. The Achievement tests are the most deviant from the regression line (Figure 1). Tests that fall below the regression line are those that show a smaller white-black difference than is predicted by their $g$ loading, and those above the regression line show larger differences than predicted.

These deviations from the regression line must mean that there are other factors besides $g$ which add to or subtract from the size of the white-black

difference in $g$. Hence, we must look at the correlation of the mean white-black $\sigma$ differences $(\bar{D})$ on each of the 24 subtests with the subtests' loadings on each of the residualized first-order factors (combined groups). We have considered only the salient loadings which help to define each factor, as the nonsalient loadings and the differences between them are neither significantly nor appreciably greater than zero and could only add "noise" to the analysis. Table 5 shows the correlation of the $g$ loading and of the salient loadings on each of the three first-order factors with the white-black differences. The correlation of factor loadings with $\bar{D}$ is negligible only for the Verbal factor, although the $g$ loadings of just the verbal tests are correlated $+.60$ with $\bar{D}$ on the verbal tests. Loadings on the Spatial factor are highly *positively* correlated $(+.845)$ with $\bar{D}$, indicating that tests of spatial ability disfavor blacks relative to whites, on average. The Memory Span factor shows the opposite, with a negative correlation $(-.825)$ with $\bar{D}$

The lower half of Table 5 shows other revealing correlations of various components of the total variance with $\bar{D}$. (Square roots of these variance components are used here to make them directly comparable with factor loadings.) Note that the specificity, error variance, and the total non-$g$ variance $(1 - g^2)$, are all *negatively* correlated with $\bar{D}$. The total variance accounted for by the residualized first-order factors is not significantly correlated $(r = +.057)$ with $\bar{D}$. This necessarily means that tests that minimize $g$ and maximize uniqueness will also minimize the white-black difference in test scores.

Is this the secret of the smaller white-black difference on the K-ABC Mental Processing (MP) IQ as compared with the WISC-R IQ? Table 6 reveals the answer. Note that the proportion of $g$ variance in the KABC-MP scales is less than in the WISC-R $(.191$ vs. $.277)$, and the specificity of the KABC-MP is

**TABLE 5**
**Correlation ($r$) between Mean White–Black**
**Differences ($\bar{D}$) on 24 Subtests of the WISC-R**
**and the K-ABC and the Subtests' Loadings**
**on Various Factors of the Hierarchical Factor**
**Analysis ($N = 172$)**

| Factor | Correlation with $\bar{D}$ |
|---|---|
| $g$ | $+.779$ |
| Verbal (V) | $-.159$ |
| Spatial (S) | $+.845$ |
| Memory Span (MS) | $-.825$ |
| $\sqrt{\text{Specificity}}$ | $-.542$ |
| $\sqrt{\text{Error Variance}}$ | $-.236$ |
| $\sqrt{\text{Communality } (h^2)}$ | $+.544$ |
| $\sqrt{(V)^2 + (S)^2 + (MS)^2}$ | $+.057$ |
| $\sqrt{1 - g^2}$ | $-.767$ |

TABLE 6
Proportions of Total Variance Associated with Various Sources for the WISC-R and K-ABC
Mental Processing (MP) and Achievement (Ach) Tests ($N$ = 172)

| Source of Variance | WISC–R | K–ABC (MP) | K–ABC (Ach) |
|---|---|---|---|
| $g$ Factor | .277 | .191 | .351 |
| Non-$g$ factors (V, S, MS) | .216 | .191 | .245 |
| Specificity | .290 | .419 | .284 |
| Error of measurement | .217 | .199 | .120 |
| Total | 1.000 | 1.000 | 1.000 |

larger than of the WISC-R (.419 vs. .290). The total non-$g$ variance of the KABC-MP is .809 as compared with .723 for the WISC-R. From this fact alone, in accord with Spearman's hypothesis, one would predict a smaller white-black difference on the K-ABC than on the WISC-R.

Finally, we computed factor scores based on the Schmid-Leiman factor analysis, and determined the white-black difference in $\sigma$ units on each of the factors. These differences are as follows: $g$, +.77; Verbal, +.20; Spatial, +.39; Memory Span, +.01. In the national standardization sample of the WISC-R, the mean white-black difference on the Full Scale IQ is 1.14 $\sigma$, and the difference in $g$ factor scores is the same (Jensen & Reynolds, 1982). When SES is partialled out of the WISC-R standardization data, however, the white-black Full Scale IQ difference is reduced to 0.80 $\sigma$ (Reynolds & Gutkin, 1981), which is close to the white-black difference of +0.77 on $g$ factor scores in the present white and black samples, which were matched on SES.

Thus, the smaller black-white difference on the K-ABC Mental Processing subtests than on the WISC-R subtests is due to two conditions: (1) the lower average $g$ loadings of the K-ABC Mental Processing subtests (*Mean g* loading = .429) than of the WISC-R subtests (*Mean g* loading = .513); and (2) the negative relation between the white-black difference and the greater specificity of the Mental Processing subtests.

It is apparent that a test constructor could manipulate the size of the black-white difference to some extent by selecting or devising subtests that include other factors besides $g$ (because $g$ is unavoidable in any cognitive test) that favor either blacks or whites. In general, it appears that, when $g$ is partialled out, tests that favor blacks over whites are those that are most loaded on a memory factor. Tests that favor whites over blacks are those most loaded (independently of $g$) on a spatial factor. The non-$g$ factors favoring either whites or blacks are more evenly balanced in the battery of WISC-R subtests than in the K-ABC battery, which contains a preponderance (10 out of 13) of subtests which favor blacks on the non-$g$ factors. If intelligence is conceived of as the general factor, or $g$, it would seem desirable that the black-white differences on other factors be bal-

anced out or minimized as much as possible. If, however, $g$ is not accepted as the construct criterion of intelligence, some explicit theoretical justification is required if intelligence (and consequently the magnitude of the black-white difference in IQ) is identified as some particular admixture of $g$ with certain other factors that are uncorrelated with $g$ and which, when summed algebraically, may favor either blacks or whites. The $g$ factor is ubiquitous in all mental tests; the non-$g$ factors are not. How can one justify any given mixture of the non-$g$ factors in a battery of tests called an intelligence test if the total score or IQ derived from the battery reflects the non-$g$ factors to such a degree that there is an appreciably less than perfect correlation between IQs and $g$ factor scores on the test?

In conclusion, the lesser black-white difference on the K-ABC than on most other IQ tests is neither a mystery nor is it attributable to any superior psychometric features of the K-ABC. It is the predictable and inevitable result of two psychometric and statistical effects manifested in the K-ABC: (1) lower $g$ loadings of the subtests, especially the Mental Processing subtests on which the IQ is based, and (2) an admixture of other factors besides $g$, primarily a memory span factor and test specificity, which, independently of $g$, favor blacks. Whether these conditions in a test are deemed desirable or undesirable depends on whether they enhance or degrade the test's construct validity as a measure of general intelligence or its practical predictive validity for criteria involving intellectual achievement.

# REFERENCES

Gordon, R.A. (1985). The black-white factor is $g$. *The Behavioral and Brain Sciences, 8*, 229–231.

Gorsuch, R.L. (1983). *Factor analysis* (2nd ed.). New York: W.B. Saunders.

Humphreys, L.G. (1985a). Race differences and the Spearman hypothesis. *Intelligence, 9*, 275–283.

Humphreys, L.G. (1985b). Attenuated hypothesis or attenuated test of hypothesis? *Intelligence, 9*, 291–295.

Jennrich, R.I. (1970). An asymptotic $\chi^2$ test for the equality of two correlation matrices. *Journal of the American Statistical Association, 65*, 904–912.

Jensen, A.R. (1980). *Bias in mental testing.* New York: The Free Press.

Jensen, A.R. (1984). The black-white difference on the K-ABC: Implications for future tests. *Journal of Special Education, 18*, 377–408.

Jensen, A.R. (1985a). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *The Behavioral and Brain Sciences, 8*, 193–219.

Jensen, A.R. (1985b). Humphreys's attenuated test of Spearman's hypothesis. *Intelligence, 9*, 285–289.

Jensen, A.R., & Reynolds, C.R. (1982). Race, social class, and ability patterns on the WISC-R. *Personality and Individual Differences, 3*, 423–438.

Jöreskog, K.G., & Sörbom, D. (1981). *LISREL V. User's guide.* Chicago, IL: National Educational Resources.

Kaufman, A.S., & Kamphaus, R.W. (1985). Factor analysis of the Kaufman Assessment Battery for Children (K-ABC) for ages 2½ through 12½ years. *Journal of Educational Psychology, 76*, 623–637.

Kaufman, A.S., & Kaufman, N.L. (1983). *Kaufman Assessment Battery for Children.* Circle Pines, MN: American Guidance Service.

Keith, T.Z., & Dunbar, S.B. (1984). Hierarchical factor analysis of the K-ABC: Testing alternate models. *Journal of Special Education, 18,* 367–375.

Naglieri, J.A. (1984). Concurrent and predictive validity of the Kaufman Assessment Battery for Children with a Navajo sample. *Journal of School Psychology, 22,* 373–380.

Naglieri, J.A. (1985a). Use of the WISC-R and K-ABC with learning disabled, borderline mentally retarded, and normal children. *Psychology in the Schools, 22,* 133–141.

Naglieri, J.A. (1985b). Assessment of mentally retarded children with the Kaufman Assessment Battery for Children. *American Journal of Mental Deficiency, 89,* 367–371.

Naglieri, J.A. (1986) *WISC-R and K-ABC comparison for matched samples of black and white children. Journal of School Psychology, 24,* 81–88.

Naglieri, J.A., & Anderson, D.F. (1985). Comparison of the WISC-R and K-ABC with gifted students. *Journal of Psychoeducational Assessment, 3,* 175–179.

Reynolds, C.R., & Gutkin, T.B. (1981). A multivariate comparison of the intellectual performance of black and white children matched on four demographic variables. *Personality and Individual Differences, 2,* 175–180.

Schmid, J., & Leiman, J.M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61.

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

Vernon, P.A. (1981). Level I and Level II: A review. *Educational Psychologist, 16,* 45–64.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children–Revised.* New York: The Psychological Corporation.

Wherry, R. (1959). Hierarchical factor solutions without rotation. *Psychometrika, 24,* 45–51.