Differences on Six Horn Abilities for 14 Age Groups Between 15–16 and 75–94 Years

Alan S. Kaufman
Psychological Assessment Resources, Inc.

James C. Kaufman
Yale University

Tsuey-Hwa Chen
American Guidance Service

Nadeen L. Kaufman
California School of Professional Psychology, San Diego

Six abilities derived from Horn’s fluid (Gf) and crystallized (Gc) theory were investigated for 1,193 individuals (607 men and 586 women), aged 15–94 years and approximately representative of the U.S. population on ethnic, socioeconomic, and geographic variables. Age differences on the Horn abilities were analyzed for separate age groups across this wide age span, both with and without control for education, using multivariate analysis of variance and multiple regression analysis. Results supported Horn’s classification of crystallized and quantitative as maintained abilities and of fluid and broad visualization as vulnerable abilities. Short-term acquisition and retrieval, usually interpreted as vulnerable, was maintained through the 60s for the present sample. The patterns of age differences on long-term retrieval tasks were a function of the type of material to be stored.

Horn (1989) and his colleagues (Horn, Donaldson, & Engstrom, 1981; Horn & Hofer, 1992) have conducted a considerable body of research on the structure of human abilities. They have identified about eight or nine factors that now define fluid (Gf) and crystallized (Gc) theory. The current theory is an expansion and refinement of the original Horn and Cattell (1966, 1967) theory that stressed the distinction between crystallized, acculturated knowledge and novel, fluid problem solving. The research that has accompanied the elucidation of the expanded Gf–Gc theory has continued to emphasize age differences on the specific ability factors, a topic of interest to the theorists from the initial development of the Gf and Gc constructs (Cattell, 1941; Cattell & Horn, 1978; Horn & Cattell, 1966, 1967).

The results of decades of research with diverse ability tests and a variety of samples have led Horn and his colleagues to distinguish between vulnerable and maintained abilities, typified by Gf and Gc, respectively (Horn, 1989; Horn & Hofer, 1992). Vulnerable abilities decline with brain damage and with age in adulthood, whereas maintained abilities do not decline with age in adulthood and return to approximate pre-injury levels following the onset of brain damage. This study investigated six abilities, three that are classified as vulnerable and three that are classified as maintained (Horn & Hofer, 1992). The vulnerable abilities are Gf, broad visualization (Gv), and short-term acquisition and retrieval (SAR); the maintained abilities are Gc, quantitative ability (Gq), and long-term retrieval (TSR).

Despite the wealth of research conducted by Horn and his coworkers, total samples have often been in the hundreds rather than the thousands; inferences about adult development, when based on relatively small samples of homogeneous age groups, are a bit tenuous. Moreover, samples have commonly comprised specific, available samples such as prisoners (e.g., Horn, 1982; Horn & Cattell, 1967) instead of carefully selected, representative samples. When large, representative samples have been used to investigate age differences in Horn’s abilities, such samples have usually comprised standardization populations for the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) or its revision, the WAIS–R (Wechsler, 1981). As Horn and Hofer (1992) noted, “Much of the evidence on vulnerable and maintained abilities has been based on the use of the WAIS” (p. 71).

Recently, additional support for the vulnerability of Gf abilities and the maintained nature of Gc abilities has been provided by cross-sectional analyses of data from new test batteries that include a variety of tasks that differ from the Wechsler mold (Kaufman & Horn, 1996; Wang & Kaufman, 1993). However, most of the data from studies of Wechsler’s scales and new tests are based on tasks that are impure from the perspective of Horn’s theory. The Wechsler Performance subtests are not categorized neatly as measuring a single Horn ability. Block design, for example, “is often regarded as a measure of Gf (although it measures a specific factor and Gv as well as Gf)” (Horn, 1989, p. 95). In a similar vein, a number of the subtests on the Gf and Gc scales that compose the Kaufman Adolescent and Adult Intelligence Test (KAIT; Kaufman & Kaufman, 1993), the instrument studied by Kaufman and Horn (1996), are not...
"pure" in the sense that they measure more than a single Horn ability. KAIT Auditory Comprehension, for example, is a Gc task that also includes SAR variance (A. Kaufman & Horn, 1996).

The primary goals of this study were to extend the existing knowledge base concerning maintained and vulnerable abilities by examining cross-sectional age differences across the life span for men and women on subtests that provide fairly pure measures of their designated Horn ability. Furthermore, this study investigated six Horn abilities; much of the previous research on maintained and vulnerable abilities has focused on the Gf and Gc abilities.

Ample evidence indicates that Gc is a maintained ability and Gf is vulnerable, on the basis of accumulated data from numerous sources: (a) cognitive investigations conducted by Horn (1982, 1985, 1989) and his colleagues (Horn et al., 1981; Horn & Hofer, 1992); (b) cross-sectional and longitudinal aging studies on Wechsler's tests (Birren & Morrison, 1961; A. Kaufman, 1990, chapter 7) and other instruments designed to measure the Horn-Cattell Gf and Gc constructs (A. Kaufman & Horn, 1996; Wang & Kaufman, 1993); and (c) neuropsychological studies of adults with unilateral brain damage (Kaufman, 1990; Turkheimer & Farace, 1992).

Evidence for the aging patterns of the other four abilities investigated in this study is less convincing. There are data to document that SAR (also referred to as Gsm) is a vulnerable ability (Horn, 1985, 1989; Horn et al., 1981). However, cross-sectional investigations of WAIS and WAIS-R Digit Span with large samples have not provided evidence of a significant decline on this subtest with increasing age (Birren & Morrison, 1961; A. Kaufman, Reynolds, & McLean, 1989). Horn and Hofer (1992) categorized Gv as a vulnerable ability but conceded, "Most research on the vulnerability of human intellect has focused on the abilities of fluid reasoning (Gf), short-term memory (Gsm), and processing speed (Gs). There is little evidence to indicate the vulnerable features of visual processing (Gv)"

(pp. 59-60).

Among maintained abilities, the evidence for Gq and TSR does not parallel the widespread evidence for Gc. For example, the maintained aspects of TSR (also called Glr or fluency of processing) does not parallel the widespread evidence for Gc. For example, most of the previous research on maintained and vulnerable abilities has focused on the Gf and Gc abilities.

Analyzes, therefore, permit verification of the claim that TSR is a maintained ability (Horn & Hofer, 1992), and it also permits a determination of whether age trends on TSR tasks are the same regardless of the nature of the stimuli that are stored and retrieved.

In addition to these two KAIT delayed recall subtests, other fairly pure measures of Horn abilities are included in two brief tests that were conormed with the KAIT on essentially the same large sample of adolescents and adults: the Kaufman Short Neuropsychological Assessment Procedure (K-SNAP: Kaufman & Kaufman, 1994b) and the Kaufman Functional Academic Skills Test (K-FAST; Kaufman & Kaufman, 1994a). The K-SNAP Composite includes three subtests that each provide measurement of a Horn ability: Number Recall (SAR), Gestalt Closure (Gv), and Four-Letter Words (Gf). The two-subtest K-FAST measures Gc (functional reading) and Gq (functional arithmetic).

The inclusion of Gestalt Closure in the analysis served additional purposes. Some researchers have disagreed with Horn's Gf-Gc interpretation of age changes on Wechsler's Verbal Scale (maintained) and Performance Scale (vulnerable). Botwinick (1977) and others (e.g., Jarvik & Bank, 1983) have interpreted the Wechsler "classical aging pattern" from a nontheoretical perspective, namely, that performance on nonspeeded tasks is maintained throughout the life span, whereas performance on speeded tasks declines. Though Gv is hypothesized by Horn to be a vulnerable ability, the Gv measure in our study (Gestalt Closure) is untimed, thereby providing a direct test of Botwinick's (1977) opposite hypothesis that nonspeeded tests are maintained throughout adulthood. Also, Gestalt Closure produced a large gender difference of about one-half SD (favoring men) in a previous investigation of data from a large portion of our sample (J. Kaufman, Chen, & A. Kaufman, 1995). That finding raised the possibility that age differences on Gv interact with gender, a hypothesis that was examined in our study.

Method

Instruments

As indicated, seven subtests were used to measure the six Horn abilities. The mean split-half reliabilities shown in parentheses for each subtest are for ages 15-16 to 75-94 years (Ns of 1,424 to 1,625; Kaufman & Kaufman, 1993, 1994a, 1994b).

Gc (knowledge or crystallized intelligence). K FAST Reading (mean reliability = .90). K-FAST Reading, a test of functional reading comprehension, is composed of 34 items. For each item, the person tries to interpret the meaning of rebus (the symbol for poison or handicapped); abbreviations ("bldg.", "non-smkrs," or "EEG"); or words, phrases, and sentences that are observed in the environment ("Push," "Exit Only," "Parental guidance suggested," or an excerpt from a catalog).

Some researchers (e.g., Bracken, 1985) have argued that conventional measures of reading are not truly measures of Gc. K-FAST Reading, however, is not a conventional test of reading achievement. Most items fit more into the category of general information (identifying rebus and abbreviations) and they have a built-in acculturation component by virtue of item context (recipes, ads, or labels). K-FAST Reading does, therefore, seem to be an appropriate measure of Gc.

Empirical evidence that K-FAST Reading measures Gc comes from a joint factor analysis with the KAIT and K-SNAP for five age groups
(total \(N = 1,270\)). Two factors emerged, defined as Gc or Gf, based on high loadings by the KAIT Crystallized and Fluid subtests. Reading had a median oblimin loading of .76 on the Gc factors versus a trivial median oblimin loading of .11 on the Gf factors (Kaufman & Kaufman, 1994a, p. 50).

Gf (broad reasoning or fluid intelligence): K-SNAP Four-Letter Words (mean reliability = .84). K-SNAP Four-Letter Words is composed of 23 items. The first item requires examinees to arrange letters into words, to provide “bottom” for the task. The remaining items require individuals to develop strategies to figure out secret letters or words by generating and evaluating hypotheses. Items 6-23 are administered in a booklet under highly speeded conditions.

Sample A illustrates the task. Examinees are shown the following stimuli on an easel page, and are told that the number in parentheses indicates the number of times the secret letter appears in that word:

```
HERE (2)
EEEK (3)
HIRE (1)
HARK (0).
```

More difficult items follow the same procedure but present more stimuli and require examinees to figure out secret words composed of two, three, or four letters.

Construct validation of Four-Letter Words as a measure of Gf is provided by a factor analytic study of 255 individuals tested on the WAIS-R, KAIT, K-FAST, and K-SNAP (A. Kaufman, Ishikuma, & Kaufman, 1994). Four-Letter Words loaded .49 on the factor identified as Gf by virtue of the loadings of .56 to .65 by the four KAIT fluid subtests: Four-Letter Words loaded below .15 on the other three factors.

Gq (quantitative thinking): K-FAST Arithmetic (mean reliability = .88). K-FAST Arithmetic is a 29-item measure of functional mathematics. For each item, the person is asked a question while looking at a visual stimulus such as a graph, the key facts to remember in a problem, a numerical symbol, a picture, a map, or the actual words to the problem. The task assesses numerical reasoning, computation skills, and mathematical concepts in real-life situations. Illustrative topics covered include adding the value of coins, understanding graphs, budgeting, and purchasing a variety of products. Items that were too schoollike and of limited practical application were minimized. Because all items are untimed, and questions are presented visually as well as orally, the role of SAR is minimized on this task.

Regarding K-FAST Arithmetic’s construct validity (Kaufman & Kaufman, 1994a), it correlated .61 to .87 (median = .80) for six samples tested on either the Stanford-Binet IV (Thorndike, Hagen, & Sattler, 1986) Quantitative Scale, WISC-R Arithmetic subtest, or K-TEA Brief (Kaufman & Kaufman, 1985) Mathematics subtest. Most of the preceding criterion variables loaded above .60 on Woodcock’s (1990) Gq dimension, along with Quantitative subtests from the Woodcock—Johnson—Revised (WJ-R) Tests of Achievement (Woodcock & Johnson, 1989). Horn and Hofer (1992) indicate that Gq “is well distinguished in the Woodcock—Johnson” (p. 69).

Gv (broad visual intelligence): K-SNAP Gestalt Closure (mean reliability = .82). K-SNAP Gestalt Closure is a 25-item subtest that requires the individual to name an object or scene pictured in a partially completed inkblot drawing; it resembles K-ABC Gestalt Closure and the Gestalt Completion test originated by Streib (1931).

Construct validation of Gestalt Closure as a measure of Gv is provided by the factor analytic study of 255 individuals (Kaufman et al., 1994). Gestalt Closure loaded .41 on the factor that was primarily defined by WAIS-R Perceptual Organization subtests (it loaded negligibly on the other three factors) and is a prototypical measure of Gv based on the writings of Horn (1989) and his colleagues (Horn & Hofer, 1992).

SAR (short-term acquisition and retrieval): K-SNAP Number Recall (mean reliability = .85). K-SNAP Number Recall is a 16-item subtest that requires the person to repeat a series of digits in the same sequence as the examiner said them; the series range in length from 2 to 9 numbers.

Construct validation of Number Recall as a measure of SAR is provided by the factor analytic study of 255 individuals (Kaufman et al., 1994). Number Recall loaded .68 on the same factor on which WAIS-R Digit Span loaded .81. Both Number Recall and Digit Span are prototypical measures of SAR (Horn & Hofer, 1992).

TSR (long-term storage and retrieval): KAIT Rebus Delayed Recall and Auditory Delayed Recall (mean reliabilities = .91 and .73, respectively). The KAIT includes a fluid test, Rebus Learning, which requires learning the word or concept associated with numerous rebus drawings and then “reading” phrases and sentences composed of these rebuses. About 45 min later, after having been administered four other subtests, they are administered Rebus Delayed Recall. This 13-item, long-term memory subtest presents new phrases and sentences for them to read that use the same symbols they were taught earlier. (They are not told that they will be retested on the symbols.)

The KAIT also includes a crystallized subtest, Auditory Comprehension, that requires listening to a recording of a mock news broadcast and then answering literal and inferential questions about each news story. About 30 min later, after having been administered three other subtests, they are administered Auditory Delayed Recall (again, without warning). In this 8-item task, individuals are asked questions about the news stories they heard earlier. The two TSR tasks, therefore, measure retrieval of information stored during either crystallized (Auditory) or fluid (Rebus) problem-solving items.

There is no empirical support for the construct validity of the delayed recall tasks, although they each possess good content validity and face validity as measures of TSR; in addition, Horn (personal communication, October 1993) considers the two delayed recall tasks to measure his TSR ability.

For both delayed recall subtests, individuals who performed poorly on the initial subtests are administered only those items on the delayed tasks to which they were exposed. The scoring systems take into account this inequity and adjust the person’s scores on each delayed task to reflect retention only in relation to the information exposed initially. The tasks, therefore, measure TSR with minimal contamination of the SAR, Gc, and Gf abilities that influenced initial performance.

Participants

The total sample for this study included 1,193 adolescents and adults (607 men and 586 women) ages 15 to 94 years (\(M = 42.4, SD = 20.7\)). The individuals in the sample included all people within the age range of 15–94 years who were administered all seven of the pertinent subtests from the K-FAST, K-SNAP, and KAIT (Kaufman & Kaufman, 1993, 1994a, 1994b). Age 15 is the youngest age for which the K-FAST is applicable; the other two batteries are also normed for 11- to 14-year-olds.

For this study, the sample was subdivided into the 14 age groups indicated in Table 1, which also presents specific background information for each age subsample and for the total group on the variables of gender, ethnic group, and educational attainment; U.S. Census percentages are also indicated.

Educational attainment equals the parents’ education for ages 15–24 years, and self-education for ages 25–94. The differences in education for the separate age groups, which generally indicate less education for the older than the younger samples, correspond approximately to cen-
greater number of women than men for most age groups above 50 years. Care was taken to ensure that all participants, especially elderly individuals, had adequate visual acuity, auditory acuity, and motor coordination for responding appropriately to test stimuli.

Results

A two-way MANOVA was conducted with age and gender as independent variables and standard scores on the seven subtests as dependent variables. Gender differences on the seven subtests were evaluated in a previous investigation (J. Kaufman et al., 1995), so the main effect for gender was not interpreted in this study. Gender was included as an independent variable to examine the Age x Gender interaction. These analyses were followed by a two-way MANCOVA with the same variables, covarying educational attainment. Parents’ education (mid-parent,
## Table 2

### Summary of MANOVA and MANCOVA Results

<table>
<thead>
<tr>
<th>Main effect/interaction</th>
<th>MANOVA</th>
<th>MANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>Wilks's lambda</td>
</tr>
<tr>
<td>Age group</td>
<td>91,7233</td>
<td>.408</td>
</tr>
<tr>
<td>Gender</td>
<td>7,1159</td>
<td>.933</td>
</tr>
<tr>
<td>Age × Gender</td>
<td>91,7233</td>
<td>.904</td>
</tr>
</tbody>
</table>

**Note.** Educational attainment was the covariate for the MANCOVAs (parents' education for ages 15-24 and self-education for ages 25-94). Standard scores on seven tests—two measuring TSR (Rebus & Auditory Delay) and one measuring Gc (Reading), Gf (Four-Letter Words), Gv (Gestalt Closure), Gq (Arithmetic), and SAR (Number Recall)—were the dependent variables in all analyses. Neither Age × Gender interaction reached significance at the .01 level. The F value for the interaction in the MANOVA was significant at the .05 level, but the alpha level was set at .01 for this study. MANOVA and MANCOVA = multivariate analysis of variance and covariance, respectively.

**p < .001.**

---

when both were known) was used for ages 15-24 and self-education was used for ages 25-94. The MANOVA and MANCOVA were followed by a series of univariate ANOVAs and ANCOVAs.

Table 2 presents summary data for the MANOVA and MANCOVA. Age and gender were significant main effects, with or without the education covariate; neither Age × Gender interaction reached significance at the alpha level set for this study (.01), although the interaction in the MANOVA did reach significance at p < .05. Effect sizes (eta²) for age (.120 in MANOVA and .112 in MANCOVA) were substantially larger than effect sizes for gender (.067 and .067) or for the interaction term (.014 and .013).

In Table 3, the main effect of age is shown to be significant for all seven subtests, both with and without the education covariate. Effect sizes are shown in Table 3 for the age main effect in the ANCOVAs (the range for the ANOVAs is from .130 for Gq to .377 for Gv). Effect sizes in the ANCOVAs exceed .20 for Gv and Gf, the value often used to denote an effect of moderate size. Age accounts for more than 20% of the variance in test scores on Gv and Gf, over and above the percentage of variance accounted for by other variables in the model.

The Age × Gender interaction is shown in Table 3 primarily because the Gv interaction was of interest in view of the large gender difference noted previously (J. Kaufman et al., 1995). As seen in Table 3, the interaction for Gv was significant in both the ANOVA and ANCOVA (effect sizes of .032 and .026, respectively). Although the interactions were also significant for Gq and TSR (Auditory Delay) in the ANOVAs, they were not significant with education covaried.

Tables 4 to 6 present means and standard deviations of the standard scores, plus the education-adjusted means, by age group, for Gc and Gf (Table 4), Gq and SAR (Table 5), and the two TSR subtests (Table 6). Table 7 presents means and adjusted means on Gv for complete age groups and separate sub-samples of men and women.

Three abilities qualify as maintained on the basis of their patterns of age differences—Gc, Gq, and TSR (Auditory)—and three appear to be vulnerable: Gf, TSR (Rebus), and Gv. SAR was not easily classifiable but is best called maintained because

---

### Table 3

<table>
<thead>
<tr>
<th>Horn ability (and test)</th>
<th>ANOVA</th>
<th>ANCOVA</th>
<th>Effect size</th>
<th>Age group</th>
<th>ANOVA</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gc (Reading)</td>
<td>24.61**</td>
<td>19.41**</td>
<td>.178</td>
<td>1.70</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Gf (Four-Letter Words)</td>
<td>36.86**</td>
<td>26.43**</td>
<td>.228</td>
<td>1.61</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>Gv (Gestalt Closure)</td>
<td>54.23**</td>
<td>42.78**</td>
<td>.323</td>
<td>2.95**</td>
<td>2.36*</td>
<td></td>
</tr>
<tr>
<td>Gq (Arithmetic)</td>
<td>13.42**</td>
<td>9.74**</td>
<td>.098</td>
<td>2.55*</td>
<td>1.48</td>
<td></td>
</tr>
<tr>
<td>SAR (Number Recall)</td>
<td>15.54**</td>
<td>8.98**</td>
<td>.091</td>
<td>1.41</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td>TSR (Rebus Delay)</td>
<td>15.02**</td>
<td>18.68**</td>
<td>.173</td>
<td>1.77</td>
<td>1.29</td>
<td></td>
</tr>
<tr>
<td>TSR (Auditory Delay)</td>
<td>23.32**</td>
<td>9.29**</td>
<td>.094</td>
<td>2.33*</td>
<td>1.34</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Educational attainment was the covariate for the ANCOVAs (parents' education for ages 15-24 and self-education for ages 25-94). Degrees of freedom for age group and Age × Gender interaction are (13, 1165) for the ANOVAs and (13, 1164) for the ANCOVAs. Effect sizes for age are for ANCOVAs. ANOVA and ANCOVA = analysis of variance and covariance, respectively.

* p < .01. ** p < .001.
no significant age differences were found between any pair of ages from 15–16 to 60–64. (Tukey's HSD test was used to follow up significant $F$ values for age group to determine which pairs of means differed significantly for the 14 age groups.)

The Tukey analysis of education-adjusted means reveals some consistencies. For the maintained abilities, ages 15–16 and usually 17–19 scored significantly lower than other groups; peak performance occurred in the 40s or 50s; no significantly lower scores occurred until ages 70–74; and markedly lower scores did not occur until ages 75–94 (who scored comparably to 15- to 16-year-olds). For the vulnerable abilities, ages 15–16 did not score significantly differently from young adults; peak performance occurred in the late teens and early 20s; the groups aged 17–24 scored significantly higher than individuals about age 40 and above; and ages 70–74 scored significantly higher than ages 75–94. For SAR, the means for ages 15–64 were about equal to each other and significantly greater than the means for ages 70+.

The significant Age × Gender interaction for Gv seems to relate to the magnitude of gender differences at different ages (Table 7). At ages 15–49, the difference favoring men is substantial (6.6 points, or .44 SD), whereas the discrepancy at ages 50–94 is trivial (1.8 points, or .12 SD).

A multiple regression analysis was also conducted to answer the same questions addressed by the multivariate and univariate analyses, and to identify the relative proportions of variance accounted for by both age and education in an adult's test scores on Horn's abilities. For this analysis, all individuals below the age of 25 were eliminated, leaving a sample size of 860. The regression analyses were restricted to ages 25–94 because of the following: (a) the adolescents' standard scores introduced curvilinearity into the relationship of age to test performance; and (b) it was desirable to conduct at least one analysis that defined education consistently for all individuals (the mix of parents' education and self-education introduced an unknown bias in the covariate analyses).

In the multiple regression analysis, age level was used to predict standard scores on each of the seven subtests. To control for education, years of education was entered first into each analysis, followed by chronological age in months. Age was considered to contribute significantly to each prediction of a Horn ability if two criteria were met: (a) significant $F$ at the .01 level, and (b) addition of at least 2% variance over and above the percentage accounted for by education alone. The .01 level was chosen because of the fact that seven analyses were conducted simultaneously. The .05 criterion was used to ensure that any increment due to age was meaningful, not just statistically significant; with a sample of 890, even small increments in $R^2$ can be significant.

The results of the multiple regression analysis are shown in Table 8. The table also includes zero-order correlations of each subtest with age and education. All correlations with age are negative, with the largest values emerging for two of the vulnerable abilities, Gv (-.566) and Gf (-.514), and the smallest for the maintained ability of Gq (-.295). Analogously, the highest correlations with education are for the three maintained abilities (about .60) and the smallest is for Gv (about .40).

Age added significantly ($p < .001$) to all multiple regression coefficients after first entering education as a control variable.

### Table 4

<table>
<thead>
<tr>
<th>Age</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Adj. M</th>
<th>Gf(Four-Letter Words) M</th>
<th>SD</th>
<th>Adj. M</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-16</td>
<td>100</td>
<td>90.9</td>
<td>12.6</td>
<td>89.1</td>
<td>103.4</td>
<td>12.8</td>
<td>102.2</td>
</tr>
<tr>
<td>17-19</td>
<td>109</td>
<td>97.8</td>
<td>10.7</td>
<td>97.3</td>
<td>106.1</td>
<td>10.2</td>
<td>105.8</td>
</tr>
<tr>
<td>20-24</td>
<td>124</td>
<td>105.7</td>
<td>10.8</td>
<td>103.9</td>
<td>108.5</td>
<td>10.7</td>
<td>107.4</td>
</tr>
<tr>
<td>25-29</td>
<td>97</td>
<td>104.4</td>
<td>10.9</td>
<td>101.9</td>
<td>105.7</td>
<td>11.0</td>
<td>104.2</td>
</tr>
<tr>
<td>30-34</td>
<td>70</td>
<td>105.7</td>
<td>11.5</td>
<td>103.8</td>
<td>104.5</td>
<td>12.2</td>
<td>103.8</td>
</tr>
<tr>
<td>35-39</td>
<td>93</td>
<td>107.8</td>
<td>10.1</td>
<td>104.3</td>
<td>104.7</td>
<td>10.3</td>
<td>102.3</td>
</tr>
<tr>
<td>40-44</td>
<td>75</td>
<td>108.6</td>
<td>11.0</td>
<td>105.8</td>
<td>103.6</td>
<td>11.2</td>
<td>102.3</td>
</tr>
<tr>
<td>45-49</td>
<td>88</td>
<td>105.1</td>
<td>13.7</td>
<td>104.8</td>
<td>103.3</td>
<td>13.5</td>
<td>102.7</td>
</tr>
<tr>
<td>50-54</td>
<td>60</td>
<td>104.4</td>
<td>12.0</td>
<td>103.4</td>
<td>100.5</td>
<td>12.3</td>
<td>99.9</td>
</tr>
<tr>
<td>55-59</td>
<td>70</td>
<td>100.1</td>
<td>15.9</td>
<td>101.4</td>
<td>98.3</td>
<td>16.3</td>
<td>99.0</td>
</tr>
<tr>
<td>60-64</td>
<td>76</td>
<td>98.0</td>
<td>14.8</td>
<td>99.7</td>
<td>92.1</td>
<td>14.5</td>
<td>93.1</td>
</tr>
<tr>
<td>65-69</td>
<td>71</td>
<td>93.7</td>
<td>8.3</td>
<td>98.0</td>
<td>91.0</td>
<td>15.3</td>
<td>93.5</td>
</tr>
<tr>
<td>70-74</td>
<td>78</td>
<td>90.5</td>
<td>8.5</td>
<td>95.0</td>
<td>87.2</td>
<td>16.9</td>
<td>90.0</td>
</tr>
<tr>
<td>75-94</td>
<td>82</td>
<td>86.2</td>
<td>16.1</td>
<td>90.0</td>
<td>80.7</td>
<td>11.9</td>
<td>82.6</td>
</tr>
</tbody>
</table>

### Table 5

<table>
<thead>
<tr>
<th>Age</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Adj. M</th>
<th>Gq (Arithmetic) M</th>
<th>SD</th>
<th>Adj. M</th>
<th>SAR (Number Recall) M</th>
<th>SD</th>
<th>Adj. M</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-16</td>
<td>100</td>
<td>94.2</td>
<td>14.1</td>
<td>92.2</td>
<td>101.1</td>
<td>12.7</td>
<td>99.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-19</td>
<td>109</td>
<td>98.5</td>
<td>13.5</td>
<td>97.8</td>
<td>102.4</td>
<td>13.8</td>
<td>101.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-24</td>
<td>124</td>
<td>105.8</td>
<td>11.3</td>
<td>103.9</td>
<td>106.4</td>
<td>13.5</td>
<td>105.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-29</td>
<td>97</td>
<td>100.9</td>
<td>15.4</td>
<td>98.5</td>
<td>104.1</td>
<td>12.4</td>
<td>102.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30-34</td>
<td>70</td>
<td>102.3</td>
<td>14.0</td>
<td>99.9</td>
<td>102.7</td>
<td>13.1</td>
<td>102.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-39</td>
<td>93</td>
<td>106.5</td>
<td>12.9</td>
<td>102.8</td>
<td>102.1</td>
<td>13.6</td>
<td>99.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-44</td>
<td>75</td>
<td>106.7</td>
<td>12.8</td>
<td>104.6</td>
<td>106.3</td>
<td>13.2</td>
<td>104.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-49</td>
<td>88</td>
<td>102.8</td>
<td>13.2</td>
<td>102.3</td>
<td>101.1</td>
<td>14.4</td>
<td>100.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-54</td>
<td>60</td>
<td>102.6</td>
<td>14.0</td>
<td>101.7</td>
<td>101.9</td>
<td>13.5</td>
<td>101.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-59</td>
<td>70</td>
<td>102.0</td>
<td>15.0</td>
<td>103.4</td>
<td>09.4</td>
<td>13.9</td>
<td>100.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60-64</td>
<td>76</td>
<td>97.9</td>
<td>14.5</td>
<td>99.6</td>
<td>97.2</td>
<td>11.8</td>
<td>98.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-69</td>
<td>71</td>
<td>97.4</td>
<td>16.5</td>
<td>101.6</td>
<td>93.6</td>
<td>15.4</td>
<td>96.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70-74</td>
<td>78</td>
<td>93.2</td>
<td>16.5</td>
<td>97.9</td>
<td>89.6</td>
<td>17.4</td>
<td>92.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75-94</td>
<td>82</td>
<td>88.0</td>
<td>14.8</td>
<td>91.8</td>
<td>86.5</td>
<td>16.3</td>
<td>88.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: All standard scores are relative to a reference group composed of all individuals ages 15–94 (N = 1,193). The differences required for significance when comparing any pair of mean standard scores, using Tukey's honestly significant difference test with a familiary alpha level of .05, are as follows (the value for the actual means is presented first, followed by the value for the education-adjusted means): Gq (7.0, 5.9) and Gf (6.6, 6.2). Adj. M = mean adjusted for the covariate, educational attainment (parents' education for ages 15–24 and self-education for ages 25–94); Gq = quantitative ability; SAR = short-term acquisition and retrieval.
Table 6
Mean Standard Scores on Two Measures of Horn's TSR Ability, Both Actual and Adjusted for Education, for 14 Age Groups

<table>
<thead>
<tr>
<th>Age</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Adj. M</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Adj. M</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-16</td>
<td>100</td>
<td>103.6</td>
<td>13.2</td>
<td>103.2</td>
<td>108.3</td>
<td>13.4</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>16-17</td>
<td>100</td>
<td>102.7</td>
<td>13.8</td>
<td>102.7</td>
<td>107.0</td>
<td>13.4</td>
<td>13.8</td>
<td></td>
</tr>
<tr>
<td>17-18</td>
<td>100</td>
<td>101.6</td>
<td>13.3</td>
<td>101.6</td>
<td>106.8</td>
<td>13.6</td>
<td>13.3</td>
<td></td>
</tr>
<tr>
<td>18-19</td>
<td>100</td>
<td>100.6</td>
<td>13.3</td>
<td>100.6</td>
<td>106.8</td>
<td>13.6</td>
<td>13.3</td>
<td></td>
</tr>
<tr>
<td>19-20</td>
<td>100</td>
<td>99.7</td>
<td>13.4</td>
<td>99.7</td>
<td>106.8</td>
<td>13.6</td>
<td>13.4</td>
<td></td>
</tr>
<tr>
<td>20-21</td>
<td>100</td>
<td>98.7</td>
<td>13.5</td>
<td>98.7</td>
<td>106.8</td>
<td>13.6</td>
<td>13.5</td>
<td></td>
</tr>
<tr>
<td>21-22</td>
<td>100</td>
<td>97.8</td>
<td>13.6</td>
<td>97.8</td>
<td>106.8</td>
<td>13.6</td>
<td>13.6</td>
<td></td>
</tr>
<tr>
<td>22-23</td>
<td>100</td>
<td>96.9</td>
<td>13.7</td>
<td>96.9</td>
<td>106.8</td>
<td>13.6</td>
<td>13.7</td>
<td></td>
</tr>
<tr>
<td>23-24</td>
<td>100</td>
<td>95.9</td>
<td>13.8</td>
<td>95.9</td>
<td>106.8</td>
<td>13.6</td>
<td>13.8</td>
<td></td>
</tr>
</tbody>
</table>

Note. All standard scores are relative to a reference group composed of all individuals ages 15-94 (N = 1,193). The differences required for significance when comparing any pair of mean standard scores, using Tukey's honestly significant difference test with a familywise alpha level of .05, are as follows (the value for the actual means is presented first, followed by the value for the education-adjusted means): TSR—Auditory (7.2, 6.4) and TSR—Rebus (6.9, 6.4). Adj. M = mean adjusted for the covariate, educational attainment (parents' education for ages 15-24 and self-education for ages 25-94). TSR = long-term retrieval.

Discussion

Fluid and Crystallized Abilities

The current findings for Gc and Gf are generally consistent with the large body of data gathered by Horn (1985, 1989) and others (A. Kaufman, 1990). However, peak performance was at ages 40-44 on Gc, lower than the peak of 55-64 for WAIS-R Verbal IQ (A. Kaufman et al., 1989) and similar peaks in Horn's (1985, 1989) research. The 40-44 peak observed in this study, however, was also seen in other cross-sectional investigations of Gf and Gc measures on which data were gathered in the late 1980s and early 1990s (A. Kaufman & Horn, 1996; Wang & Kaufman, 1993). Horn (personal communication, October 1993) has suggested that the early peak of Gc on the subtests of the KAIT (Kaufman & Kaufman, 1993), reported in the Kaufman and Horn investigation, may be related to (a) the deliberate inclusion of a Gf component in most KAIT Gc subtests to make these intellectual tasks dependent on problem solving as well as acculturation; and (b) the fact that most KAIT Gc subtests do not measure the kinds of crystallized knowledge that one obtains by intense acculturation experiences from television, newspapers, magazines, and books. Either or both of these explanations are plausible reasons for the early peak in the Gc variables in the A. Kaufman and Horn (1996) study, and also in the Wang and Kaufman (1993) investigation.

Those explanations cannot, however, account for that same finding on K-FAST Reading in this study. As indicated, this subtest is more like a general information task than like a test of reading achievement. Its items assess knowledge of rebuses, information contained on prescription drug labels, abbreviations found in recipes and want ads, and so forth. The subtest has no apparent Gf component, and its items are culture saturated with information from the media and everyday experience. The precise reasons for the earlier peak of Gc in three recent investigations compared with the later peak observed in studies of Wechsler's tests, and in Horn's own investigations, are unclear. It may be that age differences in Gc abilities are changing over time. Adults tested in the late 1980s and early 1990s may differ from adults tested in the mid-1950s and late 1970s on Wechsler's adult tests, and from other samples tested before the 1980s.

Horn's SAR Ability

In contrast to Horn's (1985, 1989) classification of SAR as a vulnerable ability, the SAR subtest in our study behaved primarily as a maintained ability, as WAIS-R Digit Span did in a previous study (A. Kaufman et al., 1989). However, these results are not necessarily inconsistent with Horn's findings that SAR is vulnerable. K-SNAP Number Recall and WAIS-R Digit Span measure primary or immediate memory, whereas Horn's SAR ability includes secondary memory and working memory as well. Much of the aging decline that Horn has noted for SAR is a function of the person's ability to concentrate (a variable that Horn, 1982, refers to as concentration on slowness or COS). Furthermore, the ability to organize information at the time of encoding (EOG; Horn, 1985) "accounts for the decline of short-term memory . . . in adulthood" (p. 282). K-SNAP Number Recall, as in Wechsler's Digits Forward, demands attention rather than concentration (A. Kaufman, 1990; Mayman, Schafer, & Rapaport, 1951) and requires little encoding because these primary memory tasks are sometimes viewed as de-emphasizing retrieval and requiring "merely immediate apprehension" (Horn et al., 1981, p. 47).

Moreover, this study investigated only one aspect of maintenance versus vulnerability, ignoring resilience to brain injury. Many studies indicate the vulnerability of Digit Span to brain injury (A. Kaufman, 1990; Matarazzo, 1972), and an initial study indicates that Number Recall is vulnerable to left hemisphere damage (Kaufman & Kaufman, 1994b, Table 7.16).

Gv and TSR Abilities

The most important findings in our study concern the Gv and TSR abilities. The Gv subtest, Gestalt Closure, proved to be a vulnerable ability, similar to the general visualization tests used in the early Horn and Cattell (1967) investigations. Gv was vulnerable for both men and women, although the Age x Gender interaction was significant. Concerning TSR, the results suggest that the vulnerability of this Horn ability may be dependent on
Table 7
Mean Standard Scores on Horn's Gv Ability (Gestalt Closure), Both Actual and Adjusted for Education, by Gender, for 14 Age Groups

<table>
<thead>
<tr>
<th>Age</th>
<th>Men</th>
<th>Women</th>
<th>Total sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>Adj. M</td>
</tr>
<tr>
<td>15-16</td>
<td>55</td>
<td>107.8</td>
<td>107.2</td>
</tr>
<tr>
<td>17-19</td>
<td>60</td>
<td>110.9</td>
<td>110.7</td>
</tr>
<tr>
<td>19-22</td>
<td>50</td>
<td>113.2</td>
<td>112.0</td>
</tr>
<tr>
<td>22-25</td>
<td>40</td>
<td>108.2</td>
<td>107.3</td>
</tr>
<tr>
<td>25-28</td>
<td>30</td>
<td>106.6</td>
<td>105.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-53</td>
<td>40</td>
<td>105.0</td>
<td>104.6</td>
</tr>
<tr>
<td>53-56</td>
<td>40</td>
<td>104.0</td>
<td>103.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70-73</td>
<td>35</td>
<td>88.5</td>
<td>87.8</td>
</tr>
<tr>
<td>73-76</td>
<td>30</td>
<td>82.3</td>
<td>82.1</td>
</tr>
<tr>
<td>76-79</td>
<td>25</td>
<td>77.8</td>
<td>77.5</td>
</tr>
<tr>
<td>79-82</td>
<td>20</td>
<td>73.3</td>
<td>73.0</td>
</tr>
</tbody>
</table>

Note: All standard scores are relative to a reference group composed of all individuals ages 15-94 (N = 1,193). The differences required for significance when comparing any pair of mean standard scores for the total age groups, using Tukey's honestly significant different test with a familywise alpha level of .05, are as follows: Gv (6.1, 5.9). Adj. M = mean adjusted for the covariate, educational attainment (parents' education for ages 15-24 and self-education for ages 25-94); Gv = broad visualization.

The vulnerability of Gv. Horn and Hofer (1992) indicated weak support for their categorization of Gv as a vulnerable ability. Our results offer additional confirmation for their classification of Gv, which produced the largest effect size of any variable studied (Table 3). Gestalt Closure is a prototypical measure of Horn's (1989) Gv factor, as it is uncontaminated by Gf, Gs, or any memory ability. For our sample, it emerged as an extremely vulnerable ability in all analyses conducted, producing a pattern of early peak and rapid decline that is associated with vulnerability (Horn, 1982, 1985, 1989). Similar patterns were observed for the Woodcock-Johnson—Revised (WJ-R; Woodcock & Johnson, 1989) Visual Processing subtests for ages 18 to 70-79 (Woodcock & Mather, 1989, Table 7-4), although the latter results are mitigated to some extent by the lack of a control for educational attainment.

Our data and those of the WJ-R suggest that Gv is vulnerable. Gestalt Closure and the WJ-R Visual Processing subtests include untimed items, arguing against Botwinick's (1977) interpretation of the classic aging pattern in terms of the speeded or nonspeeded nature of the items. As nonspeeded items, K-SNAP Gestalt Closure and WJ-R Visual Processing subtests would be hypothesized by Botwinick to be maintained abilities. Instead, they emerged as vulnerable abilities, in accordance with predictions by Horn. These findings blend with a variety of evidence from diverse sources (A. Kaufman, 1990, chapter 7) to indicate that Horn's (1985, 1989) GF-Gc theory offers

<table>
<thead>
<tr>
<th>Horn ability (and test)</th>
<th>Correlation with</th>
<th>R</th>
<th>R²</th>
<th>Increment (Age)</th>
<th>F of change</th>
<th>Educ = self-education.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gc (Reading)</td>
<td>-.418*</td>
<td>.640*</td>
<td>.676*</td>
<td>.409</td>
<td>.458</td>
<td>.049</td>
</tr>
<tr>
<td>Gf (Four-Letter Words)</td>
<td>-.314*</td>
<td>.468*</td>
<td>.500*</td>
<td>.381</td>
<td>.352</td>
<td>.034</td>
</tr>
<tr>
<td>Gv (Gestalt Closure)</td>
<td>-.366*</td>
<td>.394*</td>
<td>.427*</td>
<td>.381</td>
<td>.352</td>
<td>.034</td>
</tr>
<tr>
<td>Gq (Arithmetic)</td>
<td>-.349*</td>
<td>.617*</td>
<td>.650*</td>
<td>.500*</td>
<td>.500*</td>
<td>.034</td>
</tr>
<tr>
<td>SAR (Number Recall)</td>
<td>-.360*</td>
<td>.447*</td>
<td>.500*</td>
<td>.381</td>
<td>.352</td>
<td>.034</td>
</tr>
<tr>
<td>TSR (Rebus Delay)</td>
<td>-.424*</td>
<td>.471*</td>
<td>.550*</td>
<td>.222</td>
<td>.303</td>
<td>.081</td>
</tr>
<tr>
<td>TSR (Auditory Delay)</td>
<td>-.360*</td>
<td>.564*</td>
<td>.593*</td>
<td>.318</td>
<td>.352</td>
<td>.034</td>
</tr>
</tbody>
</table>
a more parsimonious explanation of the classic aging pattern associated with Wechsler's Verbal and Performance Scales than does Botwinick's (1977) speeded-nonspeeded interpretation.

The Gv Age X Gender interaction. Horn and Hofer (1992) indicated that performance on Gv is related to "differences in the developments of males and females" (p. 65). Jensen (1980) concluded from a review of research that "the largest and most consistently found sex difference is spatial visualization ability. . . . [The better performance by males] is not consistently established until about puberty, and it persists thereafter" (p. 626). The significant Age X Gender interaction in our study suggests that this generalization may be wrong; gender differences may not persist after age 50. Gender differences, by age, on the WAIS-R visual-spatial subtests (A. Kaufman, Kaufman-Packer, McLean, & Reynolds, 1991; A. Kaufman, McLean, & Reynolds, 1988) agree to some extent with our findings, perhaps challenging Jensen's (1980) unequivocal conclusions about visual processing.

TSR ability. The age differences on the two TSR subtests reflect separate patterns, as illustrated in Figure 1, suggesting that TSR may be inappropriately classified as a unitary factor within Horn's theory. The pattern for Auditory Delayed Recall, a measure of TSR that assesses the amount of information retained from a previous test of verbal-crystallized ability, corresponds approximately to the age pattern observed for Gc and other maintained abilities. In contrast, the pattern for Rebus Delayed Recall, a TSR test that used abstract symbols and assessed the amount of information retained from a previous test of fluid ability, corresponds to the age pattern observed for Gf and other vulnerable abilities.

This finding is intriguing because the two initial tests were given within a relatively few minutes of each other and so were the two delayed recall tests. Despite the similarities in administration procedures, adolescents at ages 15-16, for example, had much better success at retrieving rebus symbols from storage than at recalling "news" from mock broadcasts (adjusted mean score of 103.5 on the Rebus task versus 94.1 on the Auditory task). By contrast, adults ages 55-59 performed better on the Auditory (102.0) than Rebus (96.1) task. Further research on this topic is needed to clarify whether TSR is a unitary, maintained ability or whether its age patterning is a function of the stimuli to be stored and retrieved. In addition, the present tasks required storage of information for less than an hour. Because the task requirements involved retrieval following an interval of more than a few minutes, the tasks clearly fit Horn's (1989) definition of TSR. However, much of the writing about TSR by Horn and his colleagues (Horn & Hofer, 1992; Horn & Risberg, 1989) makes it clear that they are primarily interested in storage for hours, weeks, or years, such that "retrieval indicates consolidation, 'making sense out of information,' and depth of processing" (Horn & Hofer, 1992, p. 69). Our findings with the KAIT delayed recall tests, therefore, may not generalize to learning that is stored for much longer intervals.

Nonetheless, Horn (personal communication, September 1994) considers the different patterns for the two TSR tasks suggestive of "intriguing hypotheses" concerning the way that in-

Figure 1. Education-adjusted mean standard scores across the 15- to 94-year age range for two subtests that measure Horn's Long-Term Retrieval: Auditory Delayed Recall (TSR-AC) and Rebus Delayed Recall (TSR-RL).
formation becomes stored, consolidated, and retrieved. The implications of the new evidence on the TSR ability, Horn suggests, are as follows: (a) The information that gets consolidated does so through different processes, and the amount that gets consolidated relates to these processes; (b) the consolidation for elderly individuals seems to be better for information learned during Gc tasks (on which they generally perform better than on Gf tasks), with the reverse pattern holding for young adults; and (c) even though one cognitive process may be more prevalent for a specific task, processes such as encoding, making sense out of information, putting information into long-term storage, and retrieval are interrelated and cannot easily be pulled apart.

**Limitations of the Study**

The principal limitations of the investigation are the use of cross-sectional methodology and the use of only one subtest to assess five of the six Horn abilities. The cross-sectional methodology was unavoidable because that was the design of the data collection. This limitation was mitigated to some extent because (a) the age samples were fairly substantial in size relative to most cross-sectional data sets, (b) the age samples were fairly representative of the United States on important background variables, and (c) the key cohort variable of educational attainment (on which representative cross-sectional age samples differ substantially) was used as a control variable in this study. Nonetheless, other important cohort variables were uncontrolled, such as the particular history of individuals growing up during the 1929 Depression, World War II, or the Vietnam War; differences in informal educational opportunities as a function of the enlightenment of parents regarding the importance of early stimulation; differences due to the degree of accessibility of information through mass media; and so forth. This study reports age difference data; any developmental trends that are noted are purely inferential and require cross-validation with longitudinal or cross-sequential investigations. One good cross-validation procedure is the technique of cohort substitution (Kausler, 1982), that is, testing independent samples of individuals from the same cohorts at different points in time. If the samples are truly random and comparable on important variables, then cohort substitution controls for the effects of testing and experimental mortality (Schae, 1983).

The use of one subtest for most Horn abilities limits generalizability, pending cross-validation with composite measures of each Horn factor. In addition, previous factor analytic data supported the Horn label assigned to the Gc, Gf, Gv, and SAR subtests, but the labels for the Gq and TSR subtests are based on face validity and subjective judgment. Factor analytic support for all variables in this study is, therefore, essential.

One final limitation concerns the inability to analyze age differences on the Horn abilities for the separate ethnic groups that compose the sample (see Table 1) because the subsamples of minority individuals were too small to permit meaningful analysis. Such analyses would have been of interest because of the significant White–Black differences that were observed on the six Horn abilities (J. Kaufman et al., 1995). Whites outscored Blacks, with education covaried, on all subtests; differences were largest on Gq (SD = .79) and smallest on SAR (SD = .36) and Gf (SD = .31). Whereas no significant Age × Race interactions were obtained for any Horn ability (Chen, Kaufman, & Kaufman, 1994), the age groups were too broad (e.g., 55–93) to permit meaningful inferences about life span development.

**References**


Correction to Sherman et al. (1995)


In the reference list on page 444, line 36, the reference incorrectly reads, "Reitan, R. M., & Davidson, L. A." The correct spelling is Davison.