DOCUMENT RESUME

ED 389 728 TM 024 279

AUTHOR Kaufman, Alan S.; And Others

TITLE WISC-III Short Forms: Psychometric Properties vs. Clinical Relevance vs. Practical Utility.

PUB DATE 11 Nov 94

NOTE 36p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Nashville, TN, November 9-11, 1994.)

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Adolescents; Age Differences; *Children; *Clinical Diagnosis; Intelligence Quotient; *Intelligence Tests; *Psychometrics; *Test Format; Test Reliability; Test Use; Test Validity; Verbal Tests

IDENTIFIERS Performance Based Evaluation; Standardization; *Wechsler Intelligence Scale for Children III

ABSTRACT

The reliability and validity of three short forms of the Wechsler Intelligence Scale for Children III (WISC-III) were compared. Each of the short forms was a tetrad composed of two verbal and two performance subtests. The first tetrad was selected based primarily on practical considerations, particularly its brevity to administer and score. The second tetrad was selected on the basis of psychometric and clinical considerations, and the third reflected a blend of psychometric, clinical, and practical factors. WISC-III standardization data (100 males and 100 females at each year of age between 6 and 16) were used to select the short forms, to determine their psychometric properties, and to develop formulas for computing a child's estimated full-scale IQ. All three short forms had acceptable average reliability coefficients above 0.90, but the validity of the brief, practical form was too low to represent a good alternative to either of the other two tetrads. Overall, the short form that represented a compromise between psychometric, clinical, and practical variables seemed to be the best choice. (Contains 4 tables and 27 references.) (Author/SLD)

WISC-III SHORT FORMS: PSYCHOMETRIC PROPERTIES VS. CLINICAL RELEVANCE VS. PRACTICAL UTILITY

Alan S. Kaufman and Ramaswamy Balgopal
The University of Alabama

James C. Kaufman
University of Southern California

James E. McLean
The University of Alabama

Paper for presentation at
the annual meeting of the
Mid-South Educational Research Association
Nashville, TN
November 9-11, 1994

Abstract

Compared the reliability and validity of three WISC-III short forms, each a tetrad composed of two Verbal and two Performance subtests. The first tetrad was selected based primarily on practical considerations, namely its brevity to administer and score. The second was selected on the basis of psychometric and clinical considerations, and the third reflects a blend of psychometric, clinical, and practical factors. WISC-III standardization data were used to select the short forms, to determine their psychometric properties, and to develop formulas for computing a child's estimated Full Scale IQ. All three short forms had acceptable average reliability coefficients above .90, but the validity of the brief, practical short form was too low to represent a good alternative to either of the other two tetrads. Overall, the short form that represented a compromise between psychometric/ clinical and practical variables seemed to be the best choice: Similarities-Arithmetic-Picture Completion-Block Design.

WISC-III SHORT FORMS: PSYCHOMETRIC PROPERTIES VS. CLINICAL RELEVANCE VS. PRACTICAL UTILITY

Researchers have developed abbreviated versions of Wechsler's scales almost from the time that the original Wechsler-Bellevue was published more than a half-century ago (McNemar, 1950). Pioneers in constructing Wechsler short forms focused mainly on psychometric issues, specifically the validity of the abbreviated battery or the degree to which it correlated with the complete battery (Doppelt, 1956; McNemar, 1950; Silverstein, 1967). McNemar (1950) chose the 10 best pairs, triads, quartets, and quintets regardless of the scale membership (Verbal or Performance) of the chosen subtests; the only criterion was the correlation of each brief Wechsler-Bellevue battery with the Full Scale. Doppelt (1956) was concerned with scale membership, opting for the inclusion of two Verbal and two Performance subtests, but the plan for selecting the four subtests was intended as a psychometric: "It was felt that the best approach would be to select the two verbal subtests which are most highly correlated with total Verbal Score and the two Performance measures which are the best predictors of the total Performance Score" (p. 63). In fact, Doppelt (1956) used clinical considerations to break a deadlock between two comparable Verbal dyads, but it was clear that psychometric validity was the primary focus.

Kaufman (1972, 1976) stressed that when selecting short forms it is essential to give much weight to clinical factors such as ensuring that the specific abilities measured by each chosen subtest are not redundant and that the brief battery as a whole reflects an appealing clinical unit; clinical issues should not be subserved to psychometrics. Tellegen and Briggs (1967) emphasized the importance of attending to psychometric properties other than validity, most notably reliability of the chosen short form, and of providing test users with accurate formulas for converting short form sums of scaled scores to estimated IQs. Cyr and Brooker (1984) made the further suggestion that reliability be used as a joint criterion with validity when making the selection of the subtests for the short form.

Most recent developers of short forms have integrated the various psychometric and clinical suggestions about short form construction and have used methods that are far more diversified and sophisticated than McNemar's (1950) initial 10-best lists of the most valid groupings of Wechsler-Bellevue subtests (Kaufman & Applegate, 1988; Reynolds, Willson, & Clark, 1982; Silverstein, 1982). Even more recently, however, traditional notions about short form construction have been challenged by Kaufman (1990, chapter 5) and his colleagues (Grossman, Chan, Parente, & Kaufman, 1994; Grossman, Mednitsky, Dennis, Scharff, & Kaufman, 1993;

Kaufman, Ishikuma, & Kaufman, 1991). Their argument is simple: If examiners are likely to choose to administer an abbreviated battery in circumstances such as mass screening programs, when time is at a premium, then why shouldn't short forms be quick to administer and simple to score?

These researchers noted that in most previous short form investigations, one or more subtests such as Vocabulary, Picture Arrangement, Block Design, and Comprehension were chosen for psychometric and clinical reasons (Kaufman, 1972, 1976; Reynolds et al., 1982; Silverstein, 1982, 1985). All of these subtests are long to administer and the two Verbal subtests are also time-consuming and difficult to score. What may have represented sound psychometric and clinical justification was also indicative of practical folly. Why choose the cumbersome, time-consuming Picture Arrangement task when the brief, easy-to-give Picture Completion subtest is an option? Why select Comprehension, with its sometimes endless querying and frequent ambiguities in scoring, when nearly any other Verbal subtest is quicker and easier? Of what special value is the rich clinical information from Picture Arrangement and Comprehension during a short form administration? If clinical information is so important, then a short form should never be given in the first place.

Subtest administration times were computed for the WAIS-R (Wechsler, 1981) in two investigations (Ryan &

Rosenberg, 1984; Ward, Selby, & Clark, 1987). Digit Span (DSp), Arithmetic (A), Similarities (S), and Digit Symbol (DSy) averaged between 3 and 5 minutes of administration time in the two studies; Picture Completion (PC) averaged 5 to 6 minutes, and Information (I) averaged 6 to 7 minutes. In contrast, these subtests were relatively long to administer: Object Assembly (OA; 9 minutes), Comprehension (C; 9-10 minutes), Vocabulary (V; 10 minutes), Picture Arrangement (PA; 10-11 minutes), and Block Design (BD; 11 minutes). Based primarily on these administration time data, but also on some analysis of data, Kaufman et al. (1991) selected the following "amazingly" four-subtest short form of the WAIS-R: A-S-PC-DSy.

Kaufman et al. (1991) compared the brief tetrad to Silverstein's (1982, 1985) popular WAIS-R short form of A-V-PA-BD, the same four subtests chosen by Doppelt (1956) for the WAIS and by Kaufman (1976) for the WISC-R. The very short WAIS-R tetrad compared favorably in its psychometric properties when pitted against the more popular quartet. The A-V-PA-BD grouping had an average reliability coefficient of .94 and it correlated .95 with the WAIS-R Full Scale. The corresponding values for the brief A-S-PC-DSy tetrad were nearly identical: .93 and .95. Yet in contrast to the average administration time of 36 minutes for Silverstein's short form (Ryan & Rosenberg, 1984; Thompson, Howard, & Anderson, 1986; Ward et al., 1987), the A-S-PC-DSy tetrad requires about 19 minutes to administer (Grossman et al., 1993; Kaufman et al., 1991).

The average time to score the four brief subtests, compute raw scores and scaled scores, and convert the sum of scaled scores to an estimated WAIS-R Full Scale IQ was 6.3 minutes (SD = 3.1) (Grossman et al., 1993). Although comparable scoring data are not available for Silverstein's A-V-PA-BD, the inclusion of Vocabulary in that grouping means that the scoring time is undoubtedly substantially linger than 6 minutes.

For the WAIS-R, there would seem to be little justification for using the popular A-V-PA-BD quartet instead of the reliable, valid, and very short A-S-PC-DSy tetrad. The goal of this study was to ask a comparable question for the WISC-III (Wechsler, 1991): Does a WISC-III four-subtest short form that is selected solely on the basis of practical considerations (estimated administration and scoring time) provide a suitable, psychometrically sound alternative to WISC-III short forms that are selected (a) primarily on the basis of psychometric considerations, with clinical variables used as a "tie-breaker;" and (b) based on an equal blend of psychometric, clinical and practical factors?

Method

Subjects

The WISC-III standardization sample of 2,200 children and adolescents, ages 6-16 years, provided the data base for this study. The test publisher used a stratified sampling procedure and applied data gathered in 1988 by the U. S. Bureau of the Census to serve as target statistics for stratification on the variables of age, gender race/ethnicity, geographic region, and parent education. The standardization sample, composed of 100 males and 100 females at each year of age between 6 and 16, matched the Census figures with considerable precision (Wechsler, 1991).

Procedure

Selecting the short forms. The first goal of this study was to define procedures for identifying the three short forms. To identify the tetrad chosen on the basis of psychometric and clinical qualities, each Verbal dyad (every possible pair of the six Verbal subtests) was rank ordered by its: (a) split-half reliability, computed from a formula provided by Tellegen and Briggs (1967); (b) correlation with Verbal Scale; and (c) correlation with Full Scale. Also, each Performance dyad was rank ordered by its split-half reliability, correlation with Performance Scale, and correlation with Full Scale. The five regular Performance subtests plus Symbol Search were used to generate Performance dyads; Mazes was eliminated because of poor reliability, a very low "g" loading, and its generally weak psychometric properties that prompted Kaufman (1994) to urge examiners not to administer it. In selecting this short form, the best Verbal and Performance dyads, in terms of their reliability and correlations with pertinent IQ scales, were identified. Clinical factors were applied to make decisions between dyads that seemed fairly comparable in their psychometric properties.

To identify the tetrad chosen primarily on the basis of the practical quality of short administration time, data provided on the mean administration times reported for the 11 separate WAIS-R subtests (Ryan & Rosenberg, 1984; Ward et al., 1987) were applied to the WISC-III. It is likely that the administration times reported for WAIS-R subtests generalize to WISC-III subtests of the same name (or in the case of Digit Symbol/Coding, the same testing materials and procedures). Also, the administration time for Digit Symbol applies not only to Coding but to Symbol Search as well in view of the similar administration procedures. Whereas the precise mean times observed for WAIS-R subtests may be a bit different from the times that would be obtained from administration of the WISC-III, it is clear that the tests that are short or long to administer on the WAIS-R would be dichotomized in the same way on the WISC-III.

The following subtests are, therefore, considered short to administer and were eligible for inclusion on the "practical" WISC-III short form:

Verbal	Performance
Information	Picture Completion
Similarities	Coding
Arithmetic	Symbol Search
Digit Span	

It was predetermined that the practical short form would consist of two of these four Verbal subtests and two of the these three Performance subtests.

To select the third short form, the one based on a blend of psychometric, clinical, and practical considerations, the reliability and correlational data were again consulted. This time, however, practical features such as administration time and ease of scoring were used as part of the decision-making process.

Determining psychometric properties of the short forms and estimating IQs. Reliability of the three short forms for each of the 11 age groups and for the total sample was determined by using Tellegen and Briggs' (1967) formula. The validity of short forms, usually defined as the correlation of the short form with the complete Full Scale, was computed for each of the three short forms for the total sample and by age. Tellegen and Briggs' (1967) linear equating formula was used to develop conversion formulas to allow examiners to enter with the sum of the four scaled scores and convert that sum to a deviation quotient (standard score) with a mean of 100 and SD of 15. The standard error of estimate was calculated using the formula that Silverstein (1984) developed specifically for use with Tellegen and Briggs' linear equating procedure.

Results

Selecting the Three Short Forms

The psychometric/clinical tetrad. Table 1 presents the

Insert Table 1 about here

---

eight best Verbal dyads in terms of split-half reliability, correlation with Verbal Scale, and correlation with Full Scale. Table 2 presents analogous data for the eight best

---

Insert Table 2 about here

Performance dyads. The best Verbal dyad in terms of psychometric properties is S-V. For the total sample, it correlated highest with Verbal Scale, was the third-best correlate of Full Scale, and the second most reliable dyad. A-V, the dyad included in several previous short forms (Doppelt, 1956; Kaufman, 1976; Silverstein, 1982, 1985), was the best correlate of Full Scale, but was only the fifth-best correlate of Verbal Scale and the sixth most reliable. I-S and

I-V were both among the five best in each of the three categories, S-V was clearly the best in terms of psychometric properties alone. Clinically, the S-V dyad is appealing. Whereas both subtests are measures of verbal concept formation, they complement each other well in an important way: Similarities measures verbal reasoning and has a decided component of fluid intelligence, while Vocabulary depends on acquired knowledge and is a prototypical measure of crystallized ability (Horn, 1985, 1989; Kaufman, 1994). For both psychometric and clinical reasons, S-V was chosen for the short form.

Among Performance dyads, two stand out above the rest. BD-PA cracked the top three dyads for the total sample in all three psychometric categories, ranking first in its correlation with Performance Scale. BD-PC was only fourth best in its correlation with Performance Scale, but ranked first in its correlation with Full Scale and its reliability. Clinically, Picture Arrangement has more appeal than Picture Completion; it requires nonverbal reasoning, it has a crystallized as well as fluid component, and its items are excellent stimuli for evoking important emotional information about the child's preoccupations and concerns. Picture Arrangement, by virtue of its visual sequencing requirements, also matches up well with the visual-spatial Block Design.

PA-BD was chosen for the short form, just as it has been selected for many previous short forms (Doppelt, 1956; Kaufman, 1976; Silverstein, 1982, 1985). The tetrad chosen on the basis of its psychometric and clinical properties was, therefore, S-V-PA-BD. Based on data obtained for the WAIS-R (Ryan & Rosenberg, 1984; Ward et al., 1987), its estimated administration time is about 37 minutes.

The practical tetrad. Of the four short Verbal subtests listed previously, two are on the WISC-III Verbal Comprehension (VC) factor (I, S) and two comprise the Freedom from Distractibility (FD) factor (A, DSp). The choice of FD subtests is easy. Digit Span is a supplementary subtest that bears only a trivial relationship to the VC factor. In contrast, Arithmetic loads respectably on the VC factor (.56 for the total sample) when only two WISC-III factors are interpreted (Wechsler, 1991, Table 6.2) and it assesses an important array of abilities: reasoning, short and long-term memory, acquired knowledge, fluid ability (Kaufman, 1994). Arithmetic has been chosen for previous short forms that did not even consider administration time as a variable (Doppelt, 1956; Kaufman, 1972, 1976; Reynolds et al., 1982; Silverstein, 1982, 1985), and is a clear-cut choice for the WISC-III practical tetrad.

The choice between Information and Similarities is *not* as simple, and either one is a reasonable selection. Both are excellent measures of the "g" factor and of the VC factor (Kaufman, 1994). Both the I-A and S-A dyads cracked two of the three "top eights" in Table 1; I-A was more reliable (.88 vs. .87), S-A correlated higher with Verbal Scale (.92 vs. .91), and each correlated about .85 with the Full Scale.

Whereas Similarities was selected with Arithmetic as the Verbal dyad for the very brief WAIS-R tetrad, this subtest usually presents more scoring difficulties for children than for adults. Perhaps half of the WISC-III age range of 6 to 16 is in Piaget's concrete operational stage, not yet having acquired the abstract capacities that define the stage of formal operational thought. Many of the responses by children to Similarities items are concrete and functional, the kinds of responses that often must be queried and present scoring difficulties for examiners. Similarities, therefore, is still short to administer, but on the WISC-III, especially, it is neither quick nor simple to score. Information is quick to administer and score and was chosen for the WISC-III practical short form. Another advantage of choosing Information over Similarities in the present study is that it means that the psychometric/clinical tetrad of S-V-PA-BD would share no subtests in common with the practical tetrad. The lack of overlap presents an ideal situation for comparing the reliability and validity of the two short forms.

Of the three short Performance subtests, PC is the only one on the Perceptual Organization (PO) factor, and is therefore a certain choice for the practical tetrad. The remaining two brief subtests--Coding (Cd) and Symbol Search (SS)--constitute the Processing Speed (PS) factor. Although Coding is the regular subtest and Symbol Search is the supplement, Kaufman (1994) considers Symbol Search to have many clinical and psychometric advantages over Coding, and recommends that examiners routinely make the permissible substitution of Symbol Search for Coding when computing Performance and Full Scale IQs. For example, Symbol Search surpassed Coding as a correlate of the Full Scale (.56 vs. .33) and as a measure of PO when only two factors were rotated (.54 vs. .39) (Wechsler, 1991, Tables 6.2 and C.12); it also assesses mental processing speed, a cognitive variable, rather than psychomotor speed (Kaufman, 1994). Symbol Search seems to be a better choice than Coding for the short form even though the PC-SS and PC-Cd dyads are comparable psychometrically. PC-SS was the eighth-best correlate of Full Scale (.759; see Table 2), and barely missed the top eight in reliability with an overall value of .823. Corresponding values for PC-Cd are .740 and .814. However, PC-Cd correlated higher with Performance Scale (.832) than did PC-SS (.796), in large part because Coding contributes to Performance IQ and Symbol Search does not.

The practical short form, chosen for its brevity of administration time and ease of scoring, is I-A-PC-SS. This grouping includes a good measure of all four WISC-III factors: VC (Information), PO (Picture Completion), FD (Arithmetic) and PS (Symbol Search). Based on data obtained for the WAIS-R (and applying the time for DSy to SS), its estimated administration time is about 21 minutes, a savings of about 40-45% when contrasted with the psychometric/clinical short form, S-V-PA-BD.

The psychometric/clinical/practical tetrad. If practical considerations are added to the selection of a psychometrically strong and clinically rich tetrad, then the S-V-PA-BD tetrad would not be chosen. Vocabulary is a long subtest that is time-consuming to score and is impractical to include in any short form. Picture Arrangement and Block Design were the two longest WAIS-R Performance subtests to administer (Ryan & Rosenberg, 1984; Ward et al., 1987); no tetrad that weighs the practicality factor can afford the luxury of including both of these subtests.

As indicated previously, and evident from Table 2, the PC-BD dyad was just as good psychometrically as PA-BD. the latter dyad was chosen for the psychometric/clinical short form for clinical reasons. However, when practicality is included, PC-BD is a far better choice than PA-BD. Picture Arrangement is clinically superior to Picture Completion, but the latter task is much shorter and easier to administer; it also serves as a good ice-breaker and rapport-maker for any brief form, making PC-BD a sure choice for the psychometric/clinical/practical tetrad.

Despite the psychometric excellence of the S-V dyad and its clinical appeal, it is necessary to replace Vocabulary with a task that has both practical and clinical qualities. Arithmetic fits those requirements for reasons mentioned previously, and the S-A dyad has good psychometric properties: It was the fourth-best correlate of Full Scale (.852), the eighth-best correlate of Performance Scale (.915), and had an adequate split-half reliability of about .87.

The chosen tetrad is therefore S-A-PC-BD. It shares two subtests with each of the other selected short forms and represents a compromise between practicality and psychometrics. In addition to taking a shorter time to give than the 36-minute S-V-PA-BD (estimated time = 27 minutes), it is also much quicker to score, because of the substitution of Arithmetic for Vocabulary. And unlike S-V-PA-BD which assesses only the VC and PO factors, the S-A-PC-BD tetrad provides measurement of three of the four WISC-III factors.

Psychometric Properties of the Three Tetrads

Table 3 presents split-half reliability coefficients,

Insert Table 3 about here

---

standard errors of measurement (SEm), validity coefficients (correlations with Full Scale), and standard errors of estimate (SEe) for the three selected tetrads. All three short forms had reliability coefficients above .90 for the total sample, with SEm's of about 4 to 4-1/2 points. S-V-PA-BD and S-A-PC-BD each correlated .93 with the Full Scale, compared with .89 for the brief tetrad.

It is common for researchers to present adjusted validity coefficients, namely values that have been corrected for the slight spuriousness that results from the overlap in error variance that occurs when the two variables being correlated (in this case short form and complete battery) are derived from a single test administration (Silverstein, 1970). These corrections typically lower validity coefficients for Wechsler's tetrads by about .02 to .03. The formula is applicable when all subtests in a short form are included in the Full Scale, but it cannot be applied appropriately in the case of the practical short form because only three of the four subtests are included in the computation of Full Scale IQ (Symbol Search is a supplementary WISC-III subtest). When the formula is applied to the S-V-PA-BD and S-A-PC-BD tetrads the validity coefficients drop from .93 to .91 and .90, respectively; similar decreases are observed for each separate age group. The decrement for the I-A-PC-SS tetrad would probably be a bit smaller because of less spurious overlap between short form and Full Scale. Hence a reduction in the validity coefficient for the practical short form from .89 to about .87-.88 is likely.

Both the uncorrected and corrected validity coefficients have value (Kaufman, 1977; Silverstein, 1977). The uncorrected coefficients are meaningful when the short form is used to identify or screen those people who will subsequently be given the remainder of the battery; the corrected values are interpretable when the short form is used as a replacement for the complete battery (e. g., when it is administered in research investigations), such that examiners have no intention of administering the entire test battery.

For most clinical purposes, therefore, it is the actual (not the corrected) validity coefficients that are pertinent. When clinicians administer short forms, they will ordinarily administer the rest of the test battery if the profile of scaled scores on the short form or the magnitude of the estimated Full Scale IQ leads them to believe that more comprehensive assessment is needed to understand fully the individual's cognitive functioning. Hence, the actual validity coefficients were used to compute the values of SEe in Table 3 for each short form by age and for the total sample.

The average SEe is 7 points for the brief, practical short form versus about 5-1/2 points for each of the other two tetrads. The SEe provides an index of the degree to which the estimated IQ differs from the actual Full Scale IQ based on a complete administration of the WISC-III. That value is a function of the validity coefficient of each short form. In contrast, the SEm provides an index of the accuracy of the person's estimated Full Scale IQ in terms of the person's "true" estimated IQ (were it possible to administer the short form over and over without the influence of practice, fatigue, and so forth).

The SEm is a function of the reliability coefficient of each short form. The SEm of the brief, practical short form is similar in magnitude to the SEm's of the other short forms because all three tetrads have similar reliabilities. The SEe of the practical tetrad is considerably larger than the values for the other two short form because its validity coefficient is substantially lower than the coefficients for S-V-PA-BD and S-A-PC-BD. Therefore, it does less well at predicting the person's actual Full Scale IQ. When examiners administer the I-A-PC-SS short form, the odds are 2 out of 3 that the child's actual Full Scale IQ will be within 7 points of the estimated Full Scale IQ; the odds are 19 out of 20 that the actual value is within 14 points of the estimated score. With either of the other two tetrads, the corresponding values are about 5-1/2 and 11 points.

Converting Short Form Scores to Estimated Full Scale IQs

Equations for converting a person's sum of scaled scores on the four subtests to estimated Full Scale IQs were developed using Tellegen and Briggs' (1967) linear equating technique. These equations are shown here for each age group and for the total sample (Xc = sum of the four scaled scores).

Tetrad	Conversion Equation
S-V-PA-BD	Estimated Full Scale IQ = 1.6 Xc + 36 (Ages 6, 8, 10, 12-14, 16, & Total Sample)
	Estimated Full Scale IQ = 1.7 Xc + 32 (Ages 7, 9, & 11)
	Estimated Full Scale IQ = 1.5 Xc + 40 (Age 15)
S-A-PC-BD	Estimated Full Scale IQ = 1.6 Xc + 36 (Ages 6, 8-14, 16, & Total Sample)
	Estimated Full Scale IQ = 1.7 Xc + 32 (Age 7)
	Estimated Full Scale IQ = 1.5 Xc + 40 (Age 15)
I-A-PC-SS	Estimated Full Scale IQ = 1.7 Xc + 32 (Ages 6, 7, 9-14, 16, & Total Sample)
	Estimated Full Scale IQ = 1.6 Xc + 36 (Ages 8 & 15)

To use these equations, first sum the child's scaled scores on the four pertinent subtests. Enter this sum (Xc) in the equation. If examiners give the psychometric/clinical short form to a 7-year-old, for example, sum the child's scaled scores on S, V, PA, and BD. Suppose the child's sum is 43. Select the appropriate equation and enter that sum as follows:

Estimated Full Scale IQ = 1.7 (43) + 32
= 73.1 + 32 = 105.1 = 105.

Examiners may use the age-specific equations, as in the previous example, but, in actuality, the equations for the total sample for each tetrad represent a good overview for all children within the 6 to 16 year age range. These equations were used to develop Table 4, which allows examiners to

---

Insert Table 4 about here

---

enter with the sum of four scaled scores (Xc) and read off the estimated Full Scale IQ. Examiners may use this table for any of the three short forms. Since the equations for S-V-PA-BD and S-A-PC-BD are identical for the total sample, the same column is entered in Table 4 for both of these short forms.

Whenever computing a short-form estimate of Full Scale IQ, whether by formula or by entering Table 4, append the abbreviation Est. next to the value.

Discussion

The results of this study differ from the findings for a very brief WAIS-R short form (Kaufman et al., 1991). The 20-minute WAIS-R tetrad was just as valid as Silverstein's (1982, 1985) much longer A-V-PA-BD short form (.95 uncorrected correlation with Full Scale), and almost as reliable. It was, therefore, considered to be "an especially good choice for all short form uses" and "recommended for use in place of the existing brief forms of the WAIS-R" (Kaufman et al., 1991, pp. 12-13). For the WISC-III, however, the very brief practical tetrad was considerably less valid than either of the other two tetrads under investigation and was notably less efficient as a predictor of Full Scale IQ.

The lower validity of the WISC-III than WAIS-R brief tetrad is due to several factors. One concerns the inclusion of Symbol Search in the WISC-III tetrad. This subtest is not included in the Full Scale, so the overlap between the I-A-PC-SS tetrad and the Full Scale is 30% instead of the 40% for the other WISC-III tetrads, and the 36% (i. e., 4 out of 11 subtests) for the WAIS-R brief tetrad. Other things being equal, the size of the overlap affects the magnitude of the correlation with the Full Scale. (Note, also, that it is the overlap in error variance, <u>not</u> in common subtests, that is removed when Silverstein's, 1970, correction is applied.)

The overlap issue does not mean that higher validity coefficients would have been obtained if Coding had been selected for the practical tetrad in place of Symbol Search. Coding is a poor measure of "g;" even with the overlap, the PC-SS dyad correlated a bit higher with the Full Scale for the total sample (.76) than did the PC-Cd dyad (.74).

Probably the main reason that the WAIS-R brief tetrad had better psychometric properties than the WISC-III brief tetrad concerns differences in the relationships among Wechsler's subtests for children versus adults. The brief subtests correlate higher with other subtests on the WAIS-R than on the WISC-III. Picture Completion correlated .50, on the average, with other WAIS-R subtests, but only .39, on the average, with other WISC-III subtests (excluding Mazes); WAIS-R Digit Symbol averaged a correlation of .43 versus .27 for WISC-III Coding and .38 for Symbol Search (Wechsler, 1981, Table 16; Wechsler, 1991, Table C.12).

The net results of the closer relationship of Wechsler's separate abilities for adults than for children are: (a) higher "g" loadings for WAIS-R (mean = .71) than for WISC-III (.64) subtests; and (b) higher WAIS-R than WISC-III "g" loadings for the comparable brief subtests of Picture Completion (.70 vs. .60), Digit Symbol/Coding (.59 vs. .41), and Digit Span (.62 vs. .47) (Kaufman, 1990, Table 8.11; Kaufman, 1994). Consequently, brief short forms are able to predict WAIS-R Full Scale IQ with far better accuracy than they are able to predict WISC-III Full scale IQ.

A comparison of the tetrad selected on the basis of psychometric and clinical properties (S-V-PA-BD) and the one selected on the basis of psychometric, clinical and practical qualities (S-A-PC-BD) results in a dead heat. They are equally valid and about equally reliable. Yet three of the four S-A-PC-BD are of the short-to-administer variety, and only Similarities requires some subjectivity to score. Its estimated administration time of 27 minutes represents a considerable savings of time (25-30%) over the estimated 37 minutes needed for S-V-PA-BD; it is also quicker to score because it includes Arithmetic instead of Vocabulary. The estimated six or so extra minutes needed to give S-A-PC-BD instead of the practical tetrad would seem to be time well spent in terms of the payoff in added validity.

Therefore, the very brief I-A-PC-SS tetrad cannot be recommended for clinical use or for screening purposes. Its main use would probably be for research investigations or for large-scale screening programs where even the savings of six minutes per administration and several minutes of scoring time would translate to a considerable savings of time and money, a practical consideration that may be quite important. All of the short forms require validation and investigation in studies that utilize just the brief form, not the short form embedded in the complete battery. This important suggestion was made by Thompson (1987), and has been implemented for Silverstein's WAIS-R short forms (Thompson, Howard, & Anderson, 1986) and for the Kaufman et al. (1991) brief tetrad (Grossman, Chan, Parente, & Kaufman, 1994; Grossman, Mednitsky, Dennis, Scharff, & Kaufman, 1993).

As a final caution, it is essential that examiners not interpret praise for a short form's psychometric properties as a mandate for the use of that short form in place of a complete battery. That is not the case. Good short forms are valuable if their use is limited to those practical circumstances in which a complete battery, and all of valuable the psychometric and clinical information derived from a complete administration, can legitimately be bypassed in favor of a substantial reduction in testing time. Such instances include (but are not limited to) screening programs and the assessment of individuals: (a) who are "referred for psychiatric disturbance, and only a global estimate of IQ is needed in the context of a complete personality evaluation;" (b) who were "given a thorough clinical or neuropsychological evaluation within the past several years, and a quick check of current intellectual status is desired;" and (c) who do not require diagnosis of a cognitive disorder or categorization of their intelligence into a specific level of functioning for placement or other educational/vocational reasons (Kaufman, 1990, p. 127).

References

Cyr, J. J., & Brooker, B. H. (1984). Use of appropriate formulas for selecting WAIS-R short forms. Journal of Consulting and Clinical Psychology, 52, 903-905.

Doppelt, J. E. (1956). Estimating the Full Scale score on the Wechsler Adult Intelligence Scale from scores on four subtests. Journal of Consulting Psychology, 20, 63-66.

Grossman, I., Chan, T., Parente, A., & Kaufman, A. S. (1994). Validation of two new brief cognitive tests with a WAIS-R short form using a hospitalized depressed sample. Perceptual and Motor Skills, 78, 107-111.

Grossman, I., Mednitsky, S., Dennis, B., Scharff, L., & Kaufman, A. S. (1993). Validation of an "amazingly" short form of the WAIS-R for a clinically depressed sample. Journal of Psychoeducational Assessment, 11, 173-181.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), Handbook of intelligence: Theories, measurements, and applications (pp. 267-300). New York: Wiley.

Horn, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), Learning and individual differences (pp. 61-116). New York: W. H. Freeman.

Kaufman, A. S. (1972). A short form of the Wechsler Preschool and Primary Scale of Intelligence. Journal of

Consulting and Clinical Psychology, 39, 361-369.

Kaufman, A. S. (1976). A four-test short form of the WISC-R. Contemporary Educational Psychology, 1, 180-196.
Kaufman, A. S. (1977). Should short-form validity coefficients be corrected? Journal of Consulting and Clinical Psychology, 45, 1159-1161.

Kaufman, A.S. (1990). Assessing Adolescent and Adult Intelligence. Boston. MA: Allyn and Bacon.

Kaufman, A. S. (1994). Intelligent testing with the WISC-III. New York: Wiley.

Kaufman, A. S., & Applegate, B. (1988). Short forms of the K-ABC mental processing and achievement scales at ages 4 to 12-1/2 years for clinical and screening purposes. validity coefficients be corrected? Journal of Clinical Child Psychology, 17, 359-369.

Kaufman, A. S., Ishikuma, T., & Kaufman-Packer, J. L. (1991). Amazingly short forms of the WAIS-R. Journal of Psychoeducational Assessment, 9, 4-15.

McNemar, Q. (1950). On abbreviated Wechsler-Bellevue scales. Journal of Consulting Psychology, 14, 79-81.

Reynolds, C. R., Willson, V. L., & Clark, P. L. (1983). A four-test short form of the WAIS-R for clinical screening. Clinical Neuropsychology, 5, 111-116.

Ryan, J. J., & Rosenberg, S. J. (1984). Administration time estimates for WAIS-R subtests and short forms in a clinical sample. Journal of Psychoeducational Assessment, 2, 125-129.

Silverstein, A. B. (1967). Validity of WISC short forms at three age levels. Journal of Consulting Psychology, 31, 635-636.

Silverstein, A. B. (1970). Reappraisal of the validity of WAIS, WISC, and WPPSI short forms. Journal of Consulting and Clinical Psychology, 34, 12-14.

Silverstein, A. B. (1977). Comment on Kaufman's "Should short-form validity coefficients be corrected?" Journal of Consulting and Clinical Psychology, 45, 1162-1163.

Silverstein, A. B. (1982). Two- and four-subtest short forms of the Wechsler Adult Intelligence Scale - Revised. Journal of Consulting and Clinical Psychology, 50, 415-418.

Silverstein, A. B. (1984). Estimating Full Scale IQs from short forms of Wechsler's scales: Linear scaling versus linear regression. Journal of Consulting and Clinical Psychology, 52, 919.

Silverstein, A. B. (1985). Two- and four-subtest short forms of the WAIS-R: A closer look at validity and reliability. Journal of Clinical Psychology, 41, 95-97.

Tellegen, A., & Briggs, P. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. Journal of Consulting Psychology, 31, 499-506.

Thompson, A. P. (1987). Methodological issues in the clinical evaluation of two-and four-subtest short forms of the WAIS-R. Journal of Clinical Psychology, 43, 142-143.

Thompson, A. P., Howard, D., & Anderson, J. (1986). Two- and four-subtest short forms of the WAIS-R: Validity in a psychiatric sample. Canadian Journal of Behavioral Science, 18, 287-292.

Ward, L. C., Selby, R. B., & Clark, B. L. (1987). Subtest administration times and short forms of the Wechsler Adult Intelligence Scale-Revised. Journal of Clinical Psychology, 43, 276-278.

Wechsler, D. (1981). WAIS-R manual: Wechsler Adult Intelligence Scale--Revised. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). Manual for the Wechsler Intelligence Scale for Children--Third Edition (WISC-III). San Antonio, TX: The Psychological Corporation.

Table 1. The Eight Best WISC-III Verbal Dyads in Terms of Reliability, Correlation with Verbal Scale, and Correlation with Full Scale

Verbal Dyad	6	7	8	9	10	11	12	13	14	15	16	TOTAL
RELIABILITY												
Information-Vocabulary	85	86	92	89	91	92	93	93	94	94	93	915
Similarities-Vocabulary	88	87	92	89	91	91	92	89	93	92	92	905
Vocabulary-Digit Span	86	85	90	86	90	89	91	91	91	94	92	896
Information-Similarities	86	86	91	88	89	90	91	88	91	91	92	895
Vocabulary-Comprehension	85	85	92	86	90	89	91	88	90	91	88	890
Arithmetic-Vocabulary	88	83	89	84	90	89	84	90	90	91	91	886
Information-Digit Span	83	84	90	86	87	88	89	89	88	93	91	884
Information-Arithmetic	85	83	89	85	88	88	87	89	89	90	91	879
CORRELATION WITH VERBAL SCALE												
Similarities-Vocabulary	91	91	94	93	94	95	95	93	94	95	93	936
Information-Comprehension	89	91	94	92	93	93	94	93	94	96	94	934
Information-Vocabulary	90	94	95	94	93	92	94	93	93	93	93	933
Information-Similarities	90	92	93	94	94	92	94	93	93	93	94	933
Arithmetic-Vocabulary	89	91	93	92	95	93	92	92	93	95	93	931
Similarities-Comprehension	89	92	95	91	93	79	93	91	91	93	94	924
Vocabulary-Comprehension	89	90	93	91	94	91	94	93	92	94	91	921
Similarities-Arithmetic	91	87	92	89	91	92	90	90	92	93	93	915
CORRELATION WITH FULL SCALE												
Arithmetic-Vocabulary	85	83	90	84	87	90	84	86	85	89	88	862
Information-Similarities	81	83	88	86	85	85	88	84	86	86	86	856
Similarities-Vocabulary	83	81	90	84	86	91	88	85	85	88	86	854
Similarities-Arithmetic	85	81	89	84	83	84	85	85	85	86	89	852
Information-Vocabulary	82	84	90	85	86	88	86	83	87	88	84	851
Information-Comprehension	80	77	86	84	86	89	87	83	87	91	84	849
Information-Arithmetic	83	81	88	83	83	83	85	86	86	87	87	847
Similarities-Comprehension	79	79	87	82	85	88	87	83	81	87	85	840

Table 2. The Eight Best WISC-III Performance Dyads in Terms of Reliability, Correlation with Performance Scale, and Correlation with Full Scale (Decimal points omitted)

Performance Dyad	6	7	8	9	10	11	12	13	14	15	16	TOTAL
RELIABILITY												
Picture Comp-Block Design	86	87	88	88	89	87	87	88	88	92	89	882
Block Design-Symbol Search	84	84	85	84	86	87	88	88	88	91	91	872
Picture Arr-Block Design	88	86	84	84	88	82	88	88	89	88	87	869
Coding-Block Design	81	78	85	[illegible]	91	87	88	85	84	94	92	866
Block Design-Object Assem	85	81	83	87	87	83	87	89	84	90	88	863
Coding-Symbol Search	81	80	84	85	82	87	87	82	82	91	91	853
Picture Comp-Picture Arr	86	88	82	82	83	80	82	80	82	85	80	828
Coding-Picture Arr	83	83	81	82	85	79	85	79	78	86	86	824
Picture Arr-Symbol Search	83	86	79	80	80	80	84	81	82	84	84	824
CORRELATION WITH PERFORMANCE SCALE												
Picture Arr-Block Design	86	87	87	89	95	87	91	89	89	92	89	889
Picture Arr-Object Assem	90	89	88	87	91	83	90	86	85	89	89	882
Block Design-Object Assem	86	85	88	84	94	86	88	88	90	92	89	875
Picture Comp-Block Design	88	86	89	86	93	87	88	86	86	90	86	874
Picture Comp-Object Assem	88	85	84	85	93	87	87	85	85	90	85	864
Coding-Block Design	86	86	87	85	65	82	86	85	85	88	75	858
Picture Comp-Picture Arr	86	87	84	85	91	85	85	83	84	85	84	851
Coding-Object Assem	84	84	85	86	93	80	85	83	81	87	85	845
CORRELATION WITH FULL SCALE												
Picture Comp-Block Design	85	75	82	82	74	80	81	82	83	86	81	816
Picture Arr-Block Design	84	80	79	80	74	76	82	83	82	87	85	812
Block Design-Object Assem	81	75	80	79	74	78	80	80	77	86	82	792
Picture Comp-Object Assem	80	75	76	80	75	80	79	76	75	85	77	788
Picture Comp-Picture Arr	81	80	77	77	73	75	76	76	80	82	78	785
Picture Arr-Object Assem	82	82	79	77	72	72	80	75	71	84	80	785
Block Design-Symbol Search	78	73	81	77	74	70	78	74	77	82	78	766
Picture Comp-Symbol Search	78	76	79	75	73	75	73	68	76	80	76	759

Table 3. Reliability, Standard Error of Measurement, and Validity of the Three Selected Short Forms, by Age (Decimal points omitted from correlation coefficients)

	Split-Half Reliability			Standard Error of Measurement			Correlation with Full Scale			Standard Error of Estimate		
Age	SF1	SF2	SF3	SF1	SF2	SF3	SF1	SF2	SF3	SF1	SF2	SF3
6	92	92	90	4.2	4.2	4.7	92	94	89	6.0	5.2	7.0
7	91	90	89	4.5	4.7	5.0	91	91	89	6.4	6.4	7.0
8	93	93	92	4.0	4.0	5.5	92	94	90	6.0	5.2	6.7
9	91	91	89	4.5	4.5	5.0	93	93	89	5.6	5.6	7.0
10	93	92	90	4.0	4.2	4.7	91	89	87	6.4	4.2	7.6
11	91	92	91	4.5	4.2	4.5	94	91	89	5.2	6.4	7.0
12	94	92	90	3.7	4.2	4.7	94	93	89	5.2	5.6	7.0
13	93	92	90	4.0	4.2	4.7	94	92	86	5.2	6.0	7.9
14	94	92	90	3.7	4.2	4.7	93	93	90	5.6	5.6	6.7
15	94	94	93	3.7	3.7	4.0	96	94	91	4.2	5.2	6.4
16	94	93	92	3.7	4.0	4.2	94	95	90	5.2	4.7	6.7
Total	93	92	91	4.0	4.2	4.5	93	93	89	5.6	5.6	7.0

Note. SF1 = Short form 1--Similarities-Vocabulary-Picture Arrangement-Block Design;

SF2 = Short form 2--Similarities-Arithmetic-Picture Completion-Block Design;

SF3 = Short form 3--Information-Arithmetic-Picture Completion-Symbol Search.

Table 4

Conversion of the Sums of Scaled Scores on Three WISC-III Tetrads to a Deviation IQ (Standard Score) with Mean of 100 and SD of 15

	Estimated Full Scale IQ			Estimated Full Scale IQ	
Sum of Scaled Scores	S-V-PA-BD or S-A-PC-BD	I-A-PC-SS	Sum of Scaled Scores	S-V-PA-BD or S-A-PC-BD	I-A-PC-SS
76	158	160	58	129	131
75	156	159	57	127	129
74	154	158	56	126	127
73	153	156	55	124	126
72	151	154	54	122	124
71	150	153	53	121	122
70	148	151	52	119	120
69	146	149	51	118	119
68	145	148	50	116	117
67	143	146	49	114	115
66	142	144	48	113	114
65	140	142	47	111	112
64	138	141	46	110	110
63	137	139	45	108	108
62	135	137	44	106	107
61	134	136	43	105	105
60	132	134	42	103	103
59	130	132	41	102	102

(Table 4 continues)

Table 4 (continued)

Sum of Scaled Scores	Estimated Full Scale IQ S-V-PA-BD or S-A-PC-BD	Estimated Full Scale IQ I-A-PC-SS	Sum of Scaled Scores	Estimated Full Scale IQ S-V-PA-BD or S-A-PC-BD	Estimated Full Scale IQ I-A-PC-SS
40	100	100	21	70	68
39	98	98	20	68	66
38	97	97	19	66	64
37	95	95	18	65	63
36	94	93	17	63	61
35	92	92	16	62	59
34	90	90	15	60	58
33	89	88	14	58	56
32	87	86	13	57	54
31	86	85	12	55	52
30	84	83	11	54	51
29	82	81	10	52	49
28	81	80	9	50	47
27	79	78	8	49	46
26	78	76	7	47	44
25	76	74	6	46	42
24	74	73	5	44	41
23	73	71	4	42	40
22	71	69			

Note: Some values for I-A-PC-SS were smoothed slightly at the extremes.