WISC-III ADMINISTRATION, CLERICAL, AND SCORING ERRORS MADE BY STUDENT EXAMINERS

Michael Sean Belk

Certificate of Approval:

Peter Zachar, Ph. D.
Assistant Professor
Psychology

Steven G. LoBello, Ph. D.
Chairman
Associate Professor
Psychology

Glen E. Ray, Ph. D.
Associate Professor
Psychology

Roger A. Ritvo, Ph. D.
Vice Chancellor for
Academic Affairs

WISC-III ADMINISTRATION, CLERICAL, AND SCORING ERRORS MADE BY STUDENT EXAMINERS

Michael Sean Belk

A Thesis

Submitted to

The Graduate Faculty of

Auburn University Montgomery

in Partial Fulfillment of the

Requirements for the

Degree of

Masters of Science

Montgomery, Alabama

February 1, 2001

Vita

Michael Sean Belk, son of Kenneth C. Belk and Charlotte G. Belk, was born September 4, 1972 in Atlanta, Georgia. He attended the University of Georgia in Athens, Georgia and graduated with a Bachelor of Arts degree in Psychology in March, 1996. He entered Graduate School, Auburn University Montgomery, in September, 1997.

THESIS ABSTRACT

WISC-III ADMINISTRATION, CLERICAL, AND SCORING ERRORS MADE BY STUDENT EXAMINERS

Michael Sean Belk

The purpose of this study was to examine the most frequent administration, clerical, and scoring errors made by graduate student examiners who administer the WISC-III. Besides reporting the various types of errors made, a further goal was to document the effect of these errors on the obtained IQ values. Investigation of graduate students' test protocols indicated numerous administration, clerical, and scoring errors that influenced Full Scale IQ's on two-thirds of the protocols. When failure to record errors (failing to record responses on the test protocol) were omitted from the analysis, the subtests most prone to error were Comprehension, Vocabulary and Similarities. Findings of this study have implications for the education and training of graduate students enrolled in an Intelligence Assessment course.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Steven G. LoBello for his guidance and patience throughout the development and completion of this research project. His enthusiasm and wisdom were invaluable. Thanks also go to Drs. Glen E. Ray and Peter Zachar for their help in revising and fine-tuning all stages of this project. Special thanks is also given to family and friends for their support of my academic pursuits; especially my parents, Kenneth C. Belk and Charlotte G. Belk.

TABLE OF CONTENTS

WISC-III Administration, Clerical, and Scoring Errors by Student Examiners

Ever since Alfred Binet developed the first practical intelligence test in 1905, and later when the construct of intelligence quotient (IQ) was created, intelligence testing has provided valuable information about the intellectual potential of individuals and groups. Using intelligence tests, psychologists can determine if someone is of average, below average, or above average intelligence, and then make predictions about that person based on his or her IQ. For example, a number of studies report a positive correlation between IQ and academic achievement, and a negative correlation between IQ and juvenile delinquency (Barret and Dipenet 1991; Binder, 1988; see LeFrancois, 1996). Many other predictions may be made from IQ, and the social implications of a high or low IQ score may have far reaching effects for those being assessed. For example, a school psychologist may base a recommendation for college attendance on a student's IQ, or a child may be denied admittance into an honors program or a private school based on his or her IQ. Therefore, it is important to ensure that IQ test protocols are scored accurately.

Errors Made by Graduate Student Examiners

There is a large body of evidence showing that graduate students frequently make errors in the administration and scoring of Wechsler intelligence tests (Bradley, Hanna, & Lucas, 1980; Miller & Chansky, 1972; Slate & Jones, 1990b; 1990c). Slate and Jones (1990a) found that graduate students averaged slightly more than 11 errors per protocol on the Wechsler Intelligence Scale for Children-Revised (WISC-R). Slate and Chick (1989) found that graduate students averaged 15.2 total errors on each WISC-R protocol and that no graduate students turned in protocols that were free of scoring errors. While

it may not be surprising to discover that graduate students make errors when learning to administer intelligence tests, it may be surprising (and alarming) to learn that graduate students show no significant decrease in the number of errors committed over 5 to 10 practice administrations of the WISC-R (Slate & Jones 1990a). Researchers conclude that graduate student examiners continue to make errors after practice administrations because the traditional approach to teaching graduate students how to administer the WISC-R is ineffective (the traditional approach involves a practice demonstration, discussion of administration and scoring procedures, and numerous practice administrations). This approach has been criticized because teachers may not provide timely, structured feedback to the students or instruct them of problems in test administration and scoring. Moreover, it has been suggested that the traditional "practice makes perfect" method of teaching be changed because students seem to be practicing making errors and incorporating them into their routines. As a result, insufficient assessment skills have been found among graduate students and practitioners (Slate, Jones, Murray, & Coulter, 1993).

LoBello & Holley (1999) found that students continue to commit a high number of errors on the WPPSI-R, even after first learning to administer the WISC-III and the WAIS-R. The researchers found that out of 121 WPPSI-R protocols administered and scored by graduate students, none were completely free of examiner error. Moreover, examiner errors on 15% of the protocols would have resulted in IQ classification changes, with one classification changing from the Intellectually Deficient range to the Borderline range. The researchers note, "had these evaluations been conducted for actual

diagnostic purposes, several children might have been placed in inappropriate educational settings or denied admittance to certain programs" (p. 12).

A vast amount of research shows that experienced practitioners are at least as prone as graduate students to commit examiner errors (Bradley, et al., 1980; Klassen & Kishor, 1996; Slate, et al., 1993; Whitten, Slate, Jones, Shine, & Raggio, 1994), and it seems imperative to identify more efficient ways of teaching intelligence testing so that erroneous practices will not be incorporated in one's "routine." However, the high number of examiner errors committed by student examiners may be the result of a cumbersome test. The majority of errors made by graduate examiners are on the Vocabulary, Comprehension, and Similarities subtests. These findings are consistent with research on the WISC (Miller & Chansky, 1972), as well as the WISC-R (Slate & Jones, 1990a; 1990b, 1990c), and suggest that although revised, graduate students are just as likely to commit errors when administering the WISC-R as they are when they administer the WISC. It should also be noted that recent research has found that graduate students are also as likely to commit certain types of errors with the WISC-III as with the WISC-R (Klassen & Kishor, 1996).

Errors Made by Experienced Examiners

Typically the word "experience" suggests practice, wisdom, and skill in performing a particular task. One would assume, then, that those who have been trained to administer intelligence tests, and have continued to do so professionally, would be more proficient than graduate students. However, research indicates that licensed practitioners are no more likely than graduate students to properly administer and score

the WISC-R (Slate & Chick, 1989; Slate, Jones, Coulter, & Covert, 1992) and the WAIS-R (Ryan, Prifitera, & Powers, 1983; Slate, et al., 1993).

Slate, et al. (1993) found that 8 practitioners made examiner errors sufficient enough to change the IQ scores on 27 of 50 WAIS-R protocols. They also noted that these practitioners made significantly more errors in administering and scoring the WAIS-R than the graduate students in the 1990b study by Slate and Jones (the practitioners in this study had an average of 8 years professional experience and had each administered approximately 160 WAIS-R's). The practitioners committed errors on all 50 of the protocols examined in the study, and when failure to record examinee responses, circle scores, or record response times were counted as errors (as they should be), practitioners committed an average of 36.9 errors per protocol. These practitioner errors often resulted in deviations of as much as 5 points from the corrected Full Scale IQ.

Bradley, et al. (1980) found significant discrepancies among practitioners on the scoring of two "fabricated" (developed by the author) WISC-R protocols. Sixty-three members of the National Association of School Psychologists (NASP) independently scored two WISC-R protocols. On both protocols the Full Scale IQ was intended by the researchers to be 115. One of the protocols was determined by the researchers to be easy to score and the other difficult to score (the more difficult protocol contained more ambiguous responses). Scoring discrepancies were found on both the easy and the difficult protocols, and the researchers note, "Inspection of the standard deviations painfully reveals that the score an examinee receives for a given performance on WISC-R content can easily vary by 6 to 8 points" (p.531).

Klassen and Kishor (1996) performed a comparative analysis of practitioner errors on the WISC-R and the WISC-III. The purpose of their study was to determine rates of examiner's clerical errors on the WISC-III, and determine if the incidence of errors decreases as a result of experience with the WISC-III. Based on statements made in the WISC-III test manual, ("changes from the WISC-R....have made it easier for the examiner to use" and that "A final goal of WISC-III development was improvement of ...administration" (p.12) the researchers predicted that the WISC-III would be far less prone to error than the WISC-R, and that the incidence of errors would decrease with experience with the test. This was not the case, however. Klassen and Kishor found that 86% of the school psychologists who participated in their study made examiner errors on both the WISC-R and the WISC-III, with only one completing error-free protocols on both tests. In addition, 38% of the WISC-R protocols inspected contained clerical examiner errors, and 42% of the WISC-III protocols contained clerical examiner errors. The researchers also found that the number of errors committed by practitioners after using the WISC-III for 18 months did not decline significantly.

Types of Examiner Errors

Klassen and Kishor (1996) defined three types of errors that examiners tend to make on the WISC-R and the WISC-III. The first type is administration error. Administration errors are defined as straying from the prescribed, standardized procedure in giving the test. Many of these errors, such as failure to read directions verbatim, can only be detected through direct observation of the test administration, rather than protocol evaluation. However, there are other administration errors that may be detected through examination of the test protocol (i.e. failure to go in reverse sequence on the Information

subtest of the WISC-III when an examinee aged 8-16 does not obtain a perfect score on either of the first two items given). An error such as this may affect the test score, but as Thompson and Bulow (1994) point out, these types of administration errors do not always affect test scores.

The second type of error illustrated by Klassen and Kishor (1996) is the scoring error. This occurs when an examiner fails to assign the correct point value to a given response. Several subtests allow "graded" responses (0, 1, or 2 points), and assignment of a response score often requires judgment on the part of the examiner. On other subtests, Arithmetic for example, responses are either correct or incorrect, and therefore no judgment is required of the examiner. Scoring errors may affect overall test scores, however, errors that cause inflated scores on one subtest may be offset by errors that decrease scores on a different subtest. This is not to imply that scoring errors are unimportant. In fact, scoring errors could have some serious effects when assessing a client's strengths and weaknesses. Slate, et al. (1991) have shown that graduate students continue to commit scoring errors on the WAIS-R even after 10 practice administrations of the WISC-R. In the same study, they also found that practitioners commonly make scoring errors. In fact, they found that practitioners make nearly twice as many scoring errors as graduate students.

The third type of examiner error is the clerical error. Clerical errors occur when examiners make careless errors such as failing to add a column of numbers correctly or incorrectly calculating chronological age. Clerical errors also occur when examiners mistakenly add in optional subtests or use the wrong norms tables. Previous studies have found that graduate students make these errors on both the WAIS-R and the WISC-R,

and, as previously mentioned, Klassen and Kishor (1996) report that practitioners make clerical errors on the WISC-III with about the same frequency as they do on the WISC-R. Clerical errors can have serious effects on test scores, and therefore should be given special attention by the instructor. For example, if an examinee 13 years 4 months of age earned a raw score of 36 on the Vocabulary subtest, the corresponding scaled score would be 10. If the examiner incorrectly calculated the examinee's chronological age to be 14 years 4 months, the resulting scaled score would 9. This type of clerical error would not only underestimate the examinee's Verbal IQ, but also his or her Full Scale IQ and Verbal Comprehension Index. An error such as this would result in an underestimation of all subtest scaled scores, as well as all IQ and Index scores.

Causes of Examiner Errors

Previous studies of examiner error suggest that errors are caused by three factors. First, examiner errors occur during training and continue to exist afterward because the traditional training approach is inadequate (Slate & Jones, 1989; Slate, et al., 1991; Slate, et al., 1993). It also appears that many errors are due to examiner carelessness, such as circling the wrong point values or incorrectly adding points to obtain subtest totals. It has also been suggested that the scoring criteria for many of the items on the Vocabulary, Comprehension, Information, and Similarities subtests are somewhat ambiguous, making the tests difficult to score. Because these subtests require judgment on the part of the examiner, explicit scoring criteria are essential for accurate scoring. Slate and Jones (1990b) state:

> it would seem that more responses need to be included in the test manuals. If this is not possible.....supplemental listings could be provided. Perhaps more

> importantly, the conceptualization or meaning of what is acceptable as 2-, 1-, or 0-point responses should be specified more clearly. Currently test manuals permit too much examiner judgment, and, thus, subjectivity results in the scoring of individual responses. While the WISC-R manual represents an improvement of scoring criteria compared to the WISC, there still remains a substantial amount of "gray area" in which responses fall between scoring categories. As a result, scores falling in gray area would appear to require judgment and experience to assign the most appropriate point value (p.82).

Recommendations for Reducing Examiner Errors

The high number of examiner errors made by both graduate students and practitioners is alarming. It has been suggested that examiners make errors because the traditional methods of teaching intelligence testing are inadequate, and that the traditional "practice makes perfect" method of teaching needs to be changed. If this hypothesis were true, it would explain the inflated number of examiner errors committed by practitioners as well. It is also possible that examiners simply get sloppy when they get out of school and are no longer supervised.

Slate and Chick (1989) found that the most frequent errors occur on the Vocabulary, Comprehension, and Similarities subtests. They suggest that this is because these subtests require judgment on the part of the examiner, and this increases the likelihood of poor interrater reliability. The WISC-III manual reports that the interscorer agreement on the Vocabulary, Comprehension, and Similarities subtests is .98, .97, and .98, respectively (Wechsler, 1993). This finding is intriguing, especially when one considers that the most frequent errors occur on these three subtests. It may be the case

that such high interrater reliability was found with the WISC-III because only four scorers were used in the study, all of who may have been experts with the WISC-III. It may also be the case that the protocols selected were not difficult to score, containing few ambiguous responses. Whatever the case, it should be noted that the original estimates of interrater agreement found in the standardization sample are not reported in the WISC-III manual.

It has also been suggested that Wechsler test manuals contain vague and ambiguous scoring criteria, especially with regard to these three subtests, and this makes them even more difficult to score. Slate, et al. (1993) point out that although Wechsler test manuals are designed for use by competent professionals, they are also typically used as the textbook for testing courses. Because of this, the manual needs to contain more quality examples of possible responses so students will learn to properly score responses. If it is not possible to include additional responses in the test manual, supplemental listings of responses could be provided to the students.

Slate and Chick (1989) also found that examiner errors are made on the more objective subtests, such as Block Design, and that errors on the Performance subtests cannot be explained by inadequate scoring criteria. Instead, errors on the Performance subtests appear to be "mechanical" in nature and result from carelessness (i.e., circling incorrect point values). It is the researchers' conclusion that examiners need to heed past recommendations to check over protocols, at least twice and probably more, for mechanical errors. In addition, Slate and Chick suggest having paraprofessionals or clerical staff check over protocols for mathematical errors. Bradley, et al. (1980) suggest that instructors require students to go over the test protocol at least twice to check for

careless, mathematical errors. Slate and Jones (1990a; 1990b; 1990c) promote the idea of either rewarding students for properly completing test protocols, or imposing significant point deductions for careless errors.

Other efforts have been made to reduce the number of computational errors made by examiners. Thompson and Hodgins (1994) have developed the Compu-Check Form (CCF) for the purpose of checking WAIS-R clerical and computational errors. The CCF is a form that prompts examiners to double check calculations and the conversion of raw scores to scaled scores and IQ's on the WAIS-R. They found that the graduate examiners made fewer errors scoring "mock" protocols when they used the CCF. It is also interesting to note that in a field trial, 6 of 7 practitioners who used the CCF detected errors on 15 of the 47 WAIS-R protocols selected from their clinical files. The total training time for use of the CCF was 20 minutes.

With regard to the more subjectively scored subtests such as Vocabulary, Comprehension, and Similarities, it has been stated that the criteria for awarding 0, 1, or 2 points to responses must be specified more clearly, and that because these subtests require judgment, explicit scoring criteria are essential for accurate scoring (Slate and Jones, 1990a, 1990b; 1990c). Because so few examples are provided in the test manuals, it is recommended that instructors provide trainees with supplemental lists of examples indicating how to score ambiguous responses. In addition, the researchers recommend that instructors prepare several practice WAIS-R protocols for students to score, making sure to include difficult-to-score responses (Slate and Jones, 1989). By having the students grade "mock" or fabricated protocols, the instructor will be able to locate exactly where a student is making errors, and also to control the difficulty of the protocol so that

students will be prepared to score more difficult, ambiguous answers on actual administrations.

To date, only one study has been conducted analyzing examiner error on the WISC-III (Klassen & Kishor, 1996). Klassen and Kishor (1996) examined 252 test protocols (126 WISC-R and 126 WISC-III protocols) to determine whether the WISC-III was more or less prone to clerical errors than the WISC-R. They found that practitioners commit about the same number of clerical errors on the WISC-III as on the WISC-R. Eighty-six percent of the practitioners in the study committed clerical errors on the WISC-III. The incidence of error on the WISC-III is similar to that reported on the WISC-R by Wagoner (1988), although Wagoner checked for administration and scoring errors as well as clerical errors.

The present study documents the rate and type of errors made by student examiners. We examined the administration, scoring, and clerical errors made on a sample of WISC-III test protocols submitted by graduate student examiners. This study expands on the findings of Klassen and Kishor (1996), as they only looked at clerical errors, and this study looks at all types of errors. The goal is to make examiners and professors aware of pitfalls, or places where errors are common, so that both the quality of instruction, as well as the quality of practice, will be improved. It is also intended to make examiners aware that errors are common, and encourage diligence in the administration and scoring of individual intelligence tests. A final goal of this study is to determine the extent to which common errors affect test scores, and potentially, diagnoses and recommendations. The study is designed to shed light on problems associated with test administration and scoring, and to answer the following questions:

1. Which subtests on the WISC-III are most vulnerable to examiner scoring error?

2. What is the total number of errors (including failure to record errors) and adjusted errors (excluding failure to record errors)?

3. What is the effect of examiner errors on IQ scores and Index scores?

4. What is the effect of examiner errors on IQ classification?

5. What is the effect of examiner errors on Verbal-Performance differences?

Method

Participants

Twenty-one graduate students enrolled in two sections of a graduate course in intelligence testing submitted a total of 100 WISC-III protocols for evaluation. The students were in the first year of a master's degree program in applied psychology. None had prior experience in individual intelligence testing, but all had taken a course in basic psychometrics. The graduate students were 17 women and 4 men. Of the 17 women students, 12 were White, 4 were Black, and 1 was Peruvian. The 4 men students were White.

The Intelligence Testing course is structured so that students first learn to administer either the WAIS-R or the WISC-III. Half of the participants administered the WAIS-R 5 times before administering the WISC-III, and the other half administered the WISC-III first. All graduate students were asked to return their five protocols for review and evaluation as a part of this study.

Materials and Procedures

A set of checklists for evaluating the errors made by examiners on the WISC-III protocols were developed by reviewing the requirements for correct administration and scoring outlined in the WISC-III manual. A separate checklist was developed for each subtest except for the optional Mazes subtest (see Appendix A). Because the test protocols had been submitted as part of the course requirements, they had been reviewed for errors by a graduate teaching assistant who had completed the intellectual testing course. Each protocol was reviewed and re-scored using the checklists. Chronological age calculations were checked and corrections made to subtest scores and IQ values. After the test

protocols had been checked and re-scored, 5 protocols were randomly selected and reviewed a third time by a doctoral level licensed psychologist who is the regular instructor of a graduate course in IQ assessment. When scoring discrepancies were found, all unselected protocols were rechecked for like errors and corrected before randomly selecting the next 5 protocols. This process was repeated until the following criterion was met: There would be no scoring disagreements on any of the sampled protocols that affected the value of any IQ, Index, or subtest score. Following this procedure a total of 45 protocols were jointly examined until the criterion was met. Among the 45 protocols, we found a total of 217 scoring discrepancies on the test protocols, almost of all of which were easily resolved by consulting the scoring criteria in the test manual, and by discussing the proper score that should be assigned to each item. On the very few items that required more extended discussion, the final scoring decision was also made jointly, but in favor of the scorer who could present the most compelling reasons for assigning a particular score. All scoring inconsistencies were resolved to the satisfaction of both scorers before data analysis proceeded.

The guidelines for checking and re-scoring the test protocols came from the scoring criteria in the WISC-III manual (Wechsler, 1992) and the basic principles outlined by Whitten, et al. (1984). Test responses and administration procedures that violated WISC-III administration and item scoring guidelines were counted as errors. For example, if an examinee's response to a test question was assigned 2 points, but the scoring criteria in the manual indicated that the response should have been given only 1 point, a scoring error was recorded. Equivocal scoring decisions or illegible written responses were not counted as errors. Finally, when the examiner did not write down the

response given by the test participant, record the response time, or made similar omissions, these errors were counted as failure to record errors (Whitten, et al., 1994).

Results

Types of Errors

There were no protocols in our sample that were free of errors. All protocols included at least a few failure to record errors. Total errors, which include the failure to record variety, ranged from 3-116, with a mean of 45.2 errors per protocol. Adjusted errors, which exclude the failure to record type, ranged from 0-39, with a mean of 10.9 errors per protocol. In their study on the WISC-R, Slate and Chick (1989) found no protocols free from examiner error, and graduate students averaged 8.1 "independent" (adjusted) errors and 15.2 total errors per protocol.

Table 1 summarizes the frequency of various clerical and computational errors that were found on the WISC-III protocols. The frequency of such errors is low, however, miscalculating chronological age and/or using incorrect norms tables when converting raw scores to scaled scores often causes a dramatic change in IQ values.

Table 2 shows that examiners made numerous administration and scoring errors in determining subtest basal and ceiling levels, querying for additional information, and assigning points to subjects' test responses. Students were more likely to assign too many points to a response than two few. The mean number of items on a protocol that were given too many points was 2.3, whereas the mean number of items on a protocol that were given too few points was 1.07. The difference between these mean values was significant ($t = 5.2$, $p < .001$). To further analyze the nature of these scoring errors, the protocols were classified according to presence of absence of each type of error. The odds ratio was 2.3 ($\chi^2 = 7.4$, $p < .01$), indicating that students were 2.3 times more likely to assign too many points to a response, rather than too few.

Table 1

Frequency of Computational and Clerical Errors on the WISC-III

Error	Protocols n[a]	Protocols percent	Examiners n[b]	Examiners percent
Used incorrect norms tables	1	1%	1	4.8%
Raw scores copied incorrectly to front of protocol	0	0%	0	0%
Incorrect subtest scaled scores copied from tables	9	9%	6	29%
Chronological age incorrect	5	5%	4	19%
Used optional subtests in determining IQ values	9	9%	5	24%

[a] n of protocols = 100

[b] n of examiners = 21

We also looked at student examiners' understanding of the rules for following up participants' responses. Some WISC-III responses should be routinely queried, as specified in the manual. The mean number of items per protocol that should have been queried and were not was 2.8. Fewer students were prone to unnecessarily query a response, as the mean for this type of administration error was .6 per protocol. The difference between these mean values was significant ($t = 8.1$, $p < .001$). The protocols were again classified according to whether or not both errors in follow-up were present or absent. The odds ratio was 7.6 ($\chi^2 = 42.7$, $p < .01$), indicating that students were almost 8 times more likely to fail to query a response as they are to query one unnecessarily.

Table 2

Frequency of Administration and Scoring Errors on the WISC-III

Error	Errors n	Protocols n[a]	%	Examiners n[b]	%
Ceiling level incorrect	37	22	22%	14	67%
Basal level incorrect	25	18	18%	10	48%
Too many points assigned to a response	230	76	76%	20	95%
Too few points assigned to a response	107	59	59%	20	95%
Should have queried (Q) a response	280	79	79%	20	95%
Queried (Q) when not necessary	60	33	33%	18	86%

[a] n of protocols = 100

[b] n of examiners = 21

The mean number of ceiling level errors (improperly terminating a subtest) per protocol was .38 and the mean number of basal level errors (failing to establish the proper starting point for a subtest) per protocol was .25. The difference was not significant ($t = 1.3$, p = n.s.). Protocols were classified according to presence or absence of basal and ceiling level errors. The odds ratio was not significant ($\chi^2 = .76$, p = n.s.). This indicates that students were no more likely to commit errors when determining basal levels on subtests than they were to make errors in determining ceiling levels.

Table 3 presents the total number of errors with the percent of protocols that had at least one error on a given subtest. Total errors include failure to record responses or response times on protocols. The subtests are arranged in descending order with the most errors occurring on the Digit Span, Picture Completion, and Arithmetic subtests. When failure to record errors are excluded from the analysis these three subtests ranked 4th, 8th, and 9th, respectively.

Table 4 presents data for adjusted error frequencies and percent of protocols containing at least one error. This table summarizes examiner error frequencies with the failure to record error type omitted from the analysis. The subtests are listed in descending order of error frequency with the most errors occurring on the Comprehension, Vocabulary, and Similarities subtests.

Students committed about 2 to 3 scoring errors per protocol on these three subtests. It is interesting to note that the error rate drops off precipitously after these three subtests, with the average number of errors per protocol at less than 1 for each remaining subtest.

Effects of Errors on Test Scores

We also examined the effect of examiner error on the obtained IQ's and other numeric values derived from the WISC-III. When scoring errors were corrected, Full Scale IQ values were overestimated on 46 protocols (46%) and underestimated on 21 protocols (21%). However, Full Scale IQ values were unchanged on 33 (33%) of the WISC-III protocols. In addition, 29 (29%) of the Full Scale IQ values that were recalculated following correction of errors were within 1 point of the value that was originally calculated. The average change in

Table 3

Frequency of Total Examiner Errors and Percent of Protocols with Errors on WISC-III Subtests

Subtest	Mean Total Errors [a]	Percent of Protocols With Errors	Rank
Digit Span	16.3	85%	1
Picture Completion	8.1	65%	2
Arithmetic	4.5	46%	3
Vocabulary	3.1	74%	4
Comprehension	3.0	79%	5
Picture Arrangement	2.3	40%	6
Information	2.3	41%	6
Similarities	2.1	67%	7
Block Design	1.4	43%	8
Object Assembly	1.4	34%	8
Symbol Search	.6	28%	9
Coding	.4	25%	10
Total	4520		
n of Protocols	100		
Mean Errors per Protocol	45.2		
SD	24.2		
Range	3-116		

[a] Total errors includes failure to record errors.

Full Scale IQ was .83 points (range -16 to 18), well within the 3 point average standard error of measurement for the WISC-III (Wechsler, 1992). However, Full Scale IQ values on 7 protocols

changed by at least 9 points. All double digit errors in calculating Full Scale IQ were caused by either including optional subtests in the calculation of IQ values (4 protocols) or using the wrong norms tables (1 protocol).

Broad IQ classifications changed on 11 protocols. In 7 of these cases, students produced a Full Scale IQ that was higher than the recalculated value. In the 4 cases where original IQ classification was lower than the recalculated value, 1 subject originally classified as Mentally Retarded was reclassified as Borderline.

The differences between student-calculated and corrected IQ values and Index Scores were analyzed using t-tests. Table 5 gives the results of this analysis. It appears that the errors that cause differences in IQ values are generally in the Verbal section of the test. Student

examiners tend to overestimate Verbal IQ (and VCI), which in turn leads to an overestimation of Full Scale IQ.

Table 4

Frequency of Adjusted Examiner Errors and Percent of Protocols with Errors on WISC-III Subtests

Subtest	Mean Adjusted Errors [a]	Percent of Protocols With Errors	Rank
Comprehension	3	79%	1
Vocabulary	2.6	73%	2
Similarities	2	63%	3
Picture Completion	.7	29%	4
Information	.6	28%	5
Picture Arrangement	.6	24%	5
Block Design	.5	24%	6
Symbol Search	.3	15%	7
Coding	.2	11%	8
Arithmetic	.2	13%	8
Digit Span	.1	5%	9
Object Assembly	.1	5%	9
Total	1090		
n of Protocols	100		
Mean Errors per Protocol	10.9		
SD	6.7		
Range	0-39		

[a] Failure to record errors omitted

Table 5

Differences Between Student-Calculated and Corrected IQ values and Index Scores

IQ/Index Value	Mean Difference	t	p
Verbal IQ	1.03	2.75	.007
Performance IQ	.64	1.56	n.s.
Full Scale IQ	.83	2.07	.04
Verbal Comprehension Index	.76	2.92	.004
Perceptual-Organization Index	-.18	-.9	n.s.
Freedom from Distractibility Index	-.76	-1.9	n.s.
Processing Speed Index	.39	.85	n.s.
Verbal-Performance Difference	.15	.52	n.s.

Discussion

This study documents a high rate of errors made by student examiners who administer the WISC-III. The findings are consistent with Slate and Chick (1989) who analyzed the WISC-R protocols from graduate examiners. Slate and Chick (1989) found that the greatest number of errors occurred on the Vocabulary, Comprehension, and Similarities subtests. We found that when failure to record errors were not included in the analysis, the most errors were made on the Comprehension, Vocabulary, and Similarities subtests. These three subtests require the use of subjective scoring criteria for awarding points to individual items. It is likely that the amount of subjectivity involved in scoring these three subtests is at least partially responsible for the large number of errors observed in these studies.

None of the protocols evaluated in this study were free of errors, but in most cases the errors that were found had little or no effect on the obtained IQ values. However, on some protocols, the amount of change in IQ values was dramatic. Examiner errors would have resulted in IQ classification changes on 11% of the protocols, with one classification changing from the Mentally Retarded range to the Borderline range. Several children might have been placed in inappropriate educational settings or denied admittance to programs if these evaluations had been conducted for actual diagnostic purposes.

There are many other reasons to be concerned about the high rate of errors made by examiners. First, competent, ethical practitioners would want to avoid errors that might negatively influence the life of any client. Second, IQ test results often become part of legal proceedings, and a well-prepared attorney could greatly damage the

credibility of an expert witness by displaying in court a test protocol filled with errors. Third, the public may ultimately lose confidence in psychological service providers if it becomes widely known that errors in the administration and scoring of intelligence test are common.

Slate and Jones (1990a; 1990b; 1990c) have roundly criticized the WISC-R for the amount of subjectivity involved in scoring the Vocabulary, Comprehension, and Similarities subtests. In addition, they contend that there should be more examples of ambiguous responses included in the test manual and that the criteria for awarding 0, 1, or 2 points should be specified more clearly. The results of this study add quantitative support to the criticisms made by Slate and Jones about the WISC-R, and show that although revised, the WISC-III is just as prone to examiner error as the WISC-R. Also, they point to the need for a revision of test administration and scoring procedures by test developers before the next version of the WISC-III is published. Clearly, examiners (and children) would benefit from a simplified examination protocol, and one might expect to see fewer examiner errors on a less complex test.

Although the high amount of subjectivity involved in scoring the WISC-III is a source of concern, scoring ambiguity is only part of the problem. It appears that unless multiple scoring errors are committed on a protocol, their presence rarely results in broad IQ classification changes. On the other hand, the presence of only one clerical error can have a serious effect on obtained IQ scores and Index values. For example, if an examiner were to include the scaled score of an optional subtest, such as Digit Span, when determining Verbal IQ , the examinee's Verbal abilities, as well as Full Scale IQ score, might be greatly overestimated. Other examples of clerical errors which may

result in broad IQ classification changes are miscalculating the examinee's chronological age, incorrectly transferring raw scores to the front of the protocol, incorrect addition of scaled scores, and using the wrong norms tables. Thompson and Hodges (1994) found that graduate students who were trained to use their Compu-Check Form (CCF) committed far less clerical and computational errors on WAIS-R protocols than untrained graduate students. Due to the serious impact clerical errors may have on IQ scores, it seems logical for teachers of IQ assessment to consider training their students to use a similar (if not the same) checklist.

It is argued that most examiner errors could be avoided with better training (Slate, et al., 1993). Consequently, it is suggested that instructors of assessment classes re-evaluate their instructional design and consider providing more time for classroom instruction. Slate, et al. (1993) recommends that professors spend more time in the classroom going over the likely sources of administration and scoring errors, for example by presenting the students with an expanded list of examinee responses and how they should be scored. Previous research shows that testing skills on the WISC-R were improved with classroom instruction targeted toward likely sources of administration and scoring errors (Slate and Jones, 1989; Slate, et al., 1991). Conner and Woodall (1983) found that administration errors made by graduate students significantly decreased as they acquired more experience administering and scoring the WISC-R, but only after receiving structured feedback regarding the type and number of errors committed.

Slate and Jones (1989) found that giving students 2 additional hours of lecture before commencing practice administrations led to fewer errors administering and scoring the WISC-R than students who did not receive the lecture. In addition, a higher

percentage of Full Scale IQ scores were changed as a result of error for the control group. During the two-hour lecture, the errors made most frequently by the control group were described in detail, and a list of explicit rules for avoiding these errors was provided. For example, subjects were told that ambiguous verbal responses frequently caused errors, and three rules were provided to reduce this source of error. The rules were as follows: (a) "Keep in mind that you may question the examinee on unclear or vague responses that are not clear-cut 2 or 1 point answers." (b) "You have to question the examinee when (Q) follows a response in the test manual, even if it is a 0-point answer." (c) "You may not question the examinee when he or she provides a clear-cut answer that is not followed by a (Q) in the test manual.

Fantuzzo, Sisemore, and Spradlin (1983) developed the Criteria for Competent WISC-R Administration (CCWA) for the purposes of training students in psychological assessment skills. The CCWA is a "thorough and comprehensive performance checklist, which consists of 198 items distributed across 15 sections (i.e., one section for each of the 12 subtests plus an Introduction, a Conclusion, and a General Considerations sections)" (p.226). Prior to training with the CCWA, the accuracy percentages for all the students across all subsections ranged from 41% to 87%, with a mean of 60% and a standard deviation of 12.5. Following training, posttest accuracy percentages ranged from 93% to 99%, with a mean of 97% and a standard deviation of 2.06. Based on these results, the researchers deemed the CCWA an effective means of assessing and training graduate students to a pre-established competency standard of WISC-R administration. In addition, the researchers found that when using the CCWA as the central component of a training package, all participants achieved the criterion level for competency after only

three administrations (the criterion level for competency in the study was 90%). An automated version of the competency based training model for the WISC-R is also available. It too has been found to be an effective training procedure, and the training cost of this method was 50% less than the non-automated version (Blakely, Fantuzzo, & Moon, 1985).

While we await the next revision of the WISC-III, it is important that all instructors of graduate courses in intelligence testing recognize the complexity of this test. Graduate students should be oriented to the potential for committing errors in the administration and scoring, and taught specific strategies for avoiding these pitfalls. The lesson of the Conner and Woodall (1983) study, and many other similar studies, is that structured feedback may reduce certain types of errors. Anecdotally, we have found that simply telling students to record each and every response on the protocol will lead to reductions in failure to record errors. The next logical step would be to continue work on the formulation of instructional programs in the administration of the WISC-III and other individual intelligence tests that eliminates or minimizes all types of errors.

References

Blakely, W. A., Fantuzzo, J. W., & Moon, G. W. (1985). An automated competency-based model for teaching skills in the administration of the WAIS-R. Professional Psychology: Research and Practice, 16 (5), 641-647.

Bradley, F. O., Hanna, G. S., & Lucas, B. A. (1980). The reliability of scoring the WISC-R. Journal of Counseling and Clinical Psychology, 48 (4), 530-531.

Conner, R., & Woodall, F. E. (1983). The effects of experience and structured feedback on WISC-R error rates made by student-examiners. Psychology in the Schools, 20, 376-379.

Fantuzzo, J. W., Sizemore, T. A., & Spradlin, W. H. (1983). A competency-based model for teaching skills in the administration of intelligence tests. Professional Psychology: Research and Practice, 14 (2), 224-231.

Klassen, R. M., & Kishor, N. (1996). A comparative analysis of practitioners' errors on WISC-R and WISC-III. Canadian Journal of School Psychology, 12, 35-43.

LeFrancois, G. R. (1996). The lifespan (5th ed.). New York: Wadsworth.

LoBello, S. G., & Holley, G. (1999). WPPSI-R administration, clerical, and scoring errors made by student examiners. Journal of Psychoeducational Assessment, 17, 15-23.

Miller, C. K., & Chansky, N. M. (1972). Psychologists' scoring of WISC protocols. Psychology in the Schools, 9, 144-152.

Ryan, J. J., Prifitera, A., & Powers, L. (1983). Scoring reliability on the WAIS-R. Journal of Counseling and Clinical Psychology, 51 (1), 149-150.

Slate, J. R., & Chick, D. (1989), WISC-R examiner errors: Cause for concern. Psychology in the Schools, 26, 78-84.

Slate, J. R., & Jones, C. H. (1989). Can teaching of the WISC-R be improved? Quasi-experimental exploration. Professional Psychology: Research and Practice, 20 (6), 408-410.

Slate, J.R., & Jones, C. H. (1990a). Student error in administering the WISC-R: Identifying problem areas. Measurement and Evaluation in Counseling and Development, 23, 137-140.

Slate, J. R., & Jones, C. H. (1990b). Identifying students' errors in administering the WAIS-R. Psychology in the Schools, 27, 83-87.

Slate, J. R., & Jones, C. H. (1990c). Examiner errors on the WAIS-R: A source of concern. The Journal of Psychology, 124, 343-345.

Slate, J. R., Jones, C. H., & Murray, R. A. (1991). Teaching administration and scoring of the Wechsler Adult Intelligence Scale-Revised: An empirical evaluation of practice administrations. Professional Psychology: Research and Practice, 22, 375-379.

Slate , J. R., Jones, C. H., Murray, R. A., & Coulter, C. (1993). Evidence that practitioners err in administering and scoring the WAIS-R. Measurement and Evaluation in Counseling and Development, 25, 156-161.

Thompson, A. P., & Bulow, C. A. (1994). Administration error in presenting the WAIS-R blocks: Approximating the impact of scrambled presentations. Professional Psychology: Research and Practice, 25, 89-91.

Thompson, A. P., & Hodgins, C. (1994). Evaluation of checking procedure for reducing clerical and computational errors on the WAIS-R. Canadian Journal of Behavioural Science, 26, 492-504.

Wagoner, R. (1988). Scoring errors made by practicing psychologists on the WISC-R. Unpublished master's thesis, Western Carolina University.

Wechsler, D. (1992). Wechsler Intelligence Scale for Children-Third Edition. San Antonio, TX: The Psychological Corporation.

Whitten, J., Slate, J. R., Shine, A., & Raggio, D. (1994). Examiner errors in administering and scoring the WPPSI-R. Journal of Psychoeducational Assessment, 12, 49-54.

APPENDIX A

Checklist for **WISC-III** Protocols

Name of Student_____ Birth Date of Client_____ Date of Test_____

1. Chronological age: CORRECT INCORRECT

ENTER NUMBER IN BLANKS BELOW

2. _____ subtest raw scores correct, but copied incorrectly to Score Conversion page of protocol.

3. Scoring errors on individual items caused _____ subtest scaled scores to change value

4. _____ subtest scaled scores copied incorrectly from tables to front of protocol

5. _____sum of subtest scaled scores copied incorrectly from Score Conversion Page to Profile Page.

6. Addition errors:

Verbal IQ	YES	NO	Verbal Comp Index	YES	NO
Perf IQ	YES	NO	Percept Org Index	YES	NO
Full Scale IQ	YES	NO	Freed from Dist Index	YES	NO
			Process Speed Index	YES	NO

7. Wrong norms tables used YES NO

8. Used optional subtests in arriving at IQ values YES NO

****

6 . Performance IQ assigned by student _____
Recalculated Performance IQ _____
Difference (+/-) _____

7. Verbal IQ assigned by student_____
Recalculated Verbal IQ _____
Difference (+/-) _____

8. Full Scale IQ assigned by student_____
Recalculated Full Scale IQ _____
Difference (+/-) _____

9. Student V-P Difference_____

Actual V-P Difference_____

10. VCI assigned by student _____
Recalculated VCI _____
Difference (+/-) _____

11. POI assigned by student _____
Recalculated POI _____
Difference (+/-) _____

12. FFDI assigned by student_____
Recalculated WMI _____
Difference (+/-) _____

13. PSI assigned by student _____
Recalculated PSI _____
Difference _____

1. PICTURE COMPLETION

	response not recorded on protocol	0-point response given 1 point	1 point response given 0 point	Should have (Q)'d	(Q)'d when not necessary	Score not recorded
Sample: Pencil						
1. Fox (start 6-						
2. Box						
3. Cat						
4. Hand						
5. Elephant						
6. Mirror						
7. Man (start						
8. Door						
9. Ladder						
10. Clock						
11. Belt (start						
12. Woman						
13. Dice						
14. Dresser						
15. Leaf						
16. Light Bulb						
17. Piano						
18. Scissors						
19. Whistle						
20. Bathtub						
21 Thermome						
22. Male						
23. Telephone						
24. Goldfish						
25. Trellis						
26. Orange						
27. Supermkt						
28. House						
29. Umbrella						
30. Tennis						

Starting Point Correct (Depends on Age): YES NO
Credit given for items below basal level YES NO N/A
Basal level correct (2 consecutive items correct): YES NO
Reversed sequence until examinee obtained perfect scores on 2 consecutive items: YES NO N/A
Ceiling level correct (5 consecutive misses) YES NO N/A
Sum of item scores correct YES NO

2. INFORMATION

	Response not recorded	1 point response given 0 points	0 point response given 1 point	did not (Q) when appropriate	(Q)'d when should not have	Score not Recorded
1. Nose (start 6-7)						
2. Ears						
3. legs						
4. Thursday						
5. Boil (start 8-10)						
6. Coins						
7. March						
8. Week (start 11-13)						
9. Seasons						
10. Dozen						
11. Hours (14-16)						
12. Stomach						
13. Columbus						
14. Oceans						
15. Leap Year						
16. Sun						
17. Oxygen						
18. Brazil						
19. Water						
20. Bulb						
21. Population						
22. Frank						
23. Hieroglyphics						
24. Glass						
25. Greece						
26. Rust						
27. Barometer						
28. Darwin						
29. London						
30. Turpentine						

Starting Point Correct (Item #5): YES NO
Basal Level Correct (2 consecutive items): YES NO
Ceiling Level Correct (5 consecutive scores of 0) YES NO N.A.
Gave Credit for items below the basal level YES NO N.A.
Item scores summed correctly YES NO
Reversed sequence until examinee age 8-16 obtained perfect scores on
2 consecutive items (If examinee earns scores of 0 on items 5 - 12): YES NO N.A.

3. CODING

Selected Correct Coding Response Sheet:

Form A [ages 6-7]: YES NO N.A.

Form B [ages 8-16]: YES NO N.A.

Completion time recorded	YES	NO
Completion time 120" or less	YES	NO
Total Raw Score recorded	YES	NO
Total Raw Score correct	YES	NO
Practice Items Completed	YES	NO
Practice Items Counted as Part of Raw Score	YES	NO
Used black lead pencil	YES	NO

Number of scoring errors _____

4. SIMILARITIES

	Should have (Q'd)	(Q)'d when not needed	Gave 2 pts. for 1 pt response	Gave 1pt for 2 pt response	Gave any points for 0 pt. response	Gave 1 or 2 pt. response 0 points	Response not recorded	Score not recorded
Sample: Red-Blue								
1. Milk-Water								
2. Candle-Lamp								
3. Shirt-Shoe								
4. Piano-Guitar								
5. Wheel-Ball								
6. Apple-Banana*								
7. Cat-Mouse*								
8. Elbow-Knee								
9. Telephone-Radio								
10. Anger-Joy								
11. Family-Tribe								
12. Painting-Statue								
13. Ice-Steam								
14. Mountain-Lake								
15. Temp-Length								
16. First-Last								
17. Rubber-Paper								
18. 9 and 25**								
19. Salt-Water								

STARTING POINT CORRECT [started with sample, then gave item #1]: YES NO

CEILING LEVEL CORRECT [four consecutive scores of 0]: YES NO N.A.

ITEM SCORES SUMMED CORRECTLY YES NO

5. PICTURE ARRANGEMENT

	Response time not recorded on protocol	Response order not recorded on protocol	Score not circled on protocol	Incorrect score circled on protocol
1. FUN 1 (Start 6-8)				
2				
2. DOG 1				
2				
3. WALK				
4. MILK				
5. CHASE				
6. CASH				
7. WORMS				
8. SMOKE				
9. BENCH				
10. DUCK				
11. STORM				
12. WETDOG DOGWET				
13. FARM				
14. SHADOW				

Starting point correct: Item 1 (ages 6-8): YES NO NA
Item 3 (ages 9-16): YES NO NA

Counted Items 1 and 2 as failures only of both trials are failed: YES NO NA
Basal Level Correct: YES NO NA
Ceiling Level Correct (3 consecutive failures): YES NO NA
Gave credit for items below basal level YES NO NA
Sum of item scores correct YES NO

6. ARITHMETIC

	resp. not recorded	0 points for 1-point response	1 point for 0-point response	Time not recorded	Bonus pts not awarded	Bonus pts given but not earned	score not circled
1. 3 (start 6)							
2. 12							
3. 4							
4. 9							
5. 2							
6. 2 (start 7-							
7. 4							
8. 5							
9. 6							
10. 3							
11. 6							
12. 7 (start 9-							
13. 14							
14. 7							
15. $24.00							
16. 11							
17. 9							
18. 10 cents							
19. $40.00							
20. $8.50							
21. 45 mph							
22. 3/10, 6/20,							
23. $42							
24. 48							

Starting point correct (Item #1, #6, #12, or #14): YES NO
Basal level correct (2 consecutive items correct): YES NO
Credit given for items below basal level YES NO N.A.
Ceiling level correct (3 consecutive scores of 0) YES NO N.A.
Sum of item scores correct YES NO
Reversed sequences until 2 consecutive perfect scores (1 point) are obtained. (Should be done if examinee age 7-16 earns score of 0 on either of 1st two items administered) YES NO N.A.

7. BLOCK DESIGN

		time not recorded	item score not circled	second trial not given when req.	second trial given when not req.	yes/no not circled	incorrect score circled	didn't give bonus pts	bonus pts. given but not earned
1 6-7	**trial 1**								
	trial 2								
2	**trial 1**								
	trial 2								
3 8-16	**trial 1**								
	trial 2								
4									
5									
6									
7									
8									
9									
10									
11									
12									

Starting point correct (item #1 or #3): YES NO

Basal level correct (pass trial 1 of item #3): YES NO

Ceiling level correct (2 consecutive scores of 0) YES NO N/A

Sum of Item Scores correct? YES NO

Gave credit for items not administered below basal level? YES NO N/A

Reversed sequence to administer items 1 and 2 in normal sequence
(only if examinee age 8-16 receives score of 0 or 1 on item 3) YES NO N.A.

8. VOCABULARY

	should have (Q)'d	(Q)'d when not needed	gave 2 points for 1 pt. resp	gave 1 pt for 2 pt. response	gave any points for 0 pt resp	no response recorded	1 or 2 pt resp scored 0	sccore not recorded
1. Clock (6-8)								
2. Hat								
3. Umbrella (9-10)								
4. Bicycle								
5. Cow (11-13)								
6. Alphabet								
7. Donkey (14-16)								
8. Thief								
9. Leave								
10. Brave								
11. Island								
12. Ancient								
13. Nonsense								
14. Absorb								
15. Fable								
16. Precise								
17. Migrate								
18. Mimic								
19. Transparent								
20. Strenuous								
21. Boast								
22. Unanimous								
23. Seclude								
24. Rivalry								
25. Amendment								
26. Compel								
27. Affliction								
28. Imminent								
29. Aberration								
30. Dilatory								

STARTING POINT CORRECT (item #1, #3, #5, or #7 see table above): YES NO
BASAL LEVEL CORRECT (perfect 2-pt scores on 2 consecutive items): YES NO
CEILING LEVEL CORRECT (four consecutive scores of 0): YES NO N/A
SUM OF ITEM SCORES CORRECT YES NO
GAVE CREDIT FOR ITEMS BELOW BASAL: YES NO NA
Reversed sequence until examinee earned perfect scores (2-points) on 2 consecutive items (should be done if examinee earns scores of 0 or 1 on either of first two items administered): YES NO N.A.

9. OBJECT ASSEMBLY

	response time not recorded	number of correct junctures not recorded	score not circled	incorrect score circled	bonus pts. earned but not given	bonus pts. given but not earned
1. Girl start all						
2. Car						
3. Horse						
4. Ball						
5. Face						

STARTING POINT CORRECT (Item #1): YES NO
ADMINISTERED ALL ITEMS: YES NO
TOTAL SCORE CORRECT: YES NO

10. COMPREHENSION

	should have (Q)'d	(Q)'d when not needed	gave 2 points for 1 point response	gave 1 point for 2 point response	gave any points for 0 point resp	no response recorded	1 or 2 pt. resp given 0 points
1. Cut Finger (start							
2. Smoke*							
3. Seatbelts							
4. Find Wallet							
5. Lose Ball							
6. Lights*							
7. Rules*							
8. Fight							
9. Telephone Book							
10. Inspect meat							
11. License plates*							
12. Newspaper*							
13. Secret Ballot							
14. Stamps							
15. Paperback*							
16. Promise							
17. Senators*							
18. Freedom							

*REQUIRES TWO 1-POINT RESPONSES. DID EXAMINER (Q) FOR ADDITIONAL RESPONSE TO

ITEM #2	YES	NO	N/A	ITEM #6	YES	NO	N/A
ITEM #7	YES	NO	N/A	ITEM #11	YES	NO	N/A
ITEM #12	YES	NO	N/A	ITEM #15	YES	NO	N/A
ITEM #17	YES	NO	N/A	ITEM #18	YES	NO	N/A

Starting Point Correct: (Item #1 for all ages): YES NO

CEILING LEVEL CORRECT (3 consecutive scores of 0) YES NO N/A

ITEM SCORES SUMMED CORRECTLY YES NO

11. SYMBOL SEARCH

1. Correct form used:			
Part A (Age 6-7)	**YES**	**NO**	**NA**
Part B (Age 8-16)	**YES**	**NO**	**NA**
2. Completion time recorded	**YES**	**NO**	
3. Completion time 120" or less	**YES**	**NO**	
4. Number correct recorded	**YES**	**NO**	
5. Number correct wrong	**YES**	**NO**	
6. Number incorrect recorded	**YES**	**NO**	
7. Number incorrect wrong	**YES**	**NO**	
8. Total Raw Score recorded	**YES**	**NO**	
9. Total Raw Score wrong	**YES**	**NO**	
10. Used black lead pencil	**YES**	**NO**	
11. Sample items completed	**YES**	**NO**	
12. Practice items completed	**YES**	**NO**	

12. DIGIT SPAN
Digits Forward

		Response not recorded	Trial score not recorded	Item score not recorded	Incorrect trial score	Incorrect item score
1	2-9					
	4-6					
2	3-8-6					
	6-1-2					
3	3-4-1-7					
	6-1-5-8					
4	8-4-2-3-9					
	5-2-1-8-6					
5	3-8-9-1-7-4					
	7-9-6-4-8-3					
6	5-1-7-4-2-3-8					
	9-8-5-2-1-6-3					
7	1-6-4-5-9-7-6-3					
	2-9-7-6-3-1-5-4					
8	5-3-8-7-1-2-4-					
	4-2-6-9-1-7-8-					

Starting point correct (Item #1): YES NO
Administered both trials of each item: YES NO
Ceiling level correct (score of 0 on both trials of any item) YES NO N/A
Sum of Digits Forward item scores correct YES NO

12. DIGIT SPAN
Digits Backward

		Response not recorded	Trial score not recorded	Item score not recorded	Trial score incorrect	Item score incorrect
1.	2-5					
	6-3					
2.	5-7-4					
	2-5-9					
3.	7-2-9-6					
	8-4-9-3					
4.	4-1-3-5-7					
	9-7-8-5-2					
5.	1-6-5-2-9-8					
	3-6-7-1-9-4					
6.	8-5-9-2-3-4-2					
	4-5-7-9-2-8-1					
7.	6-9-1-6-3-2-5-8					
	3-1-7-9-5-4-8-2					

Starting point correct (Item #1): YES NO
Administered both trials of each item YES NO
Ceiling level correct (Score of 0 on both trials of any item): YES NO N/A
Sum of Digits Backward item scores correct: YES NO
Sum of Digits Forward+Digits Backward correct (Max= 30): YES NO

WISC-III ERROR TABLE

	Total Number of Errors	Errors Not Including Failure to Record
Picture Completion		
Information		
Coding		
Similarities		
Picture Arrangement		
Arithmetic		
Block Design		
Vocabulary		
Object Assembly		
Comprehension		
Symbol Search		
Digit Span		

Totals _____ _____

RECALCULATE IQ VALUES:

Performance	**Verbal**	**POI**	**VCI**
Pic Comp _____	Inform_____	Pic Comp_____	Inform_____
Coding _____	Sim _____	Pic Arr _____	Sim _____
Pic Arr _____	Ari _____	Block _____	Voc _____
Block _____	Voc _____	OA _____	Comp _____
OA _____	Comp_____	**TOTAL** _____	**TOTAL**_____
TOTAL _____	**TOTAL** _____	**POI** _____	**VCI** _____
PIQ _____	**VIQ** _____		