EXAMINER ERROR IN ADMINISTERING THE WECHSLER SCALES OF INTELLIGENCE – "PRACTICE MAKES PERFECT" – OR DOES IT?

Gregory A. Marty, Captain, USAF

Certificate of Approval:

Peter Zachar, Ph. D.
Associate Professor
Psychology

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

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

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

EXAMINER ERROR IN ADMINISTERING THE WECHSLER SCALES OF INTELLIGENCE – "PRACTICE MAKES PERFECT" – OR DOES IT?

Gregory A. Marty, Captain, USAF

A Thesis Proposal

Submitted to

The Graduate Faculty

Of Auburn University Montgomery

In Partial Fulfillment of the

Requirements for the

Degree of

Master of Science

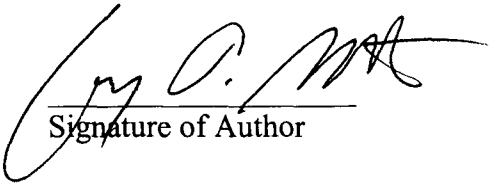
Montgomery, Alabama

April 4, 2002

EXAMINER ERROR IN ADMINISTERING THE WECHSLER SCALES OF INTELLIGENCE – "PRACTICE MAKES PERFECT" – OR DOES IT?

Gregory Alan Marty

Permission is granted to Auburn University at Montgomery to make copies of this thesis at its discretion, upon the request of individuals or institutions and at their expense. The author reserves all publication rights.



Signature of Author

9 MAY 02

Date

Copy sent to:

Name Date

VITA

Gregory Alan Marty, son of David and Alice Marty was born August 17, 1968 in Beloit Wisconsin. Gregory enlisted in the United States Air Force in May of 1992. He graduated from the University of Maryland-University College with a Bachelor of Science Degree in Psychology in August, 1995. He attended Air Force Officer Training School in March of 1996 and was commissioned a Second Lieutenant June 21st 1996. He entered graduate school at Auburn University Montgomery in August, 2000.

THESIS ABSTRACT

EXAMINER ERROR IN ADMINISTERING THE WECHSLER SCALES OF INTELLIGENCE – "PRACTICE MAKES PERFECT" – OR DOES IT?

Gregory Alan Marty

This study examined the effects of practice administrations without structured feedback on the number of errors made by graduate student examiners who administer the WAIS-III. Data were collected from 14 students in four individual intelligence testing classes taught between winter 1998 and winter 2000. 56 WAIS-III protocols were rescored using a checklist built using the Wechsler Scoring manual. Total errors were counted as well as *failure to record* errors (failure to record responses on the protocol). In the absence of structured feedback, additional practice administrations did not reduce the average amount of total errors or *failure to record* errors made by student examiners.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Steven G. LoBello for his patience, friendship, and confidence in me to complete this project. Thanks for helping me see the light at the end of the tunnel and helping me realize it wasn't a train. Thanks also to Drs. Glen Ray and Peter Zachar for inspiring me to want to become a smarter person and better student. Special thanks to my parents Dave and Alice Marty for supporting me and instilling in me the ethics to want to succeed, and the morals to do it the right way. Finally, thank you Fletcher, Linda, and Miller Majors for making me feel at home in Montgomery.

TABLE OF CONTENTS

Examiner Error in Administering the Wechsler Scales of Intelligence

"Practice makes Perfect" - Or Does it?

Richard M. Thorndike once used the analogy that "Psychological tests are a little like firearms; used carefully by someone with appropriate training, they can be very valuable, but in the hands of the incautious or inexpert, they are potentially dangerous" (as cited in Aiken, 1996). But what exactly does "dangerous" mean in the world of psychological testing? To be sure, there is potential danger in the misdiagnosis of psychological disorders due to inexpert or careless administration of psychological tests. In our nation's schools, psychological tests are used to make educational placement decisions that have lifelong impact on children and families. These as well as other important uses for intelligence tests make it easy to see why it is imperative that the techniques used for training examiners be scrutinized. Furthermore, because of decreasing budgets and increased demands on time, these teaching methods must be as economical and efficient as possible (Moon, Fantuzzo & Gorsuch, 1986).

One of the most frequently used psychological tests is the individual intelligence test. Among intelligence tests, the Wechsler scales of intelligence are the most frequently used tests across all disciplines and settings (Archer, Maruish, Imhof & Pitrowski, 1991; Lees-Haley, 1992; Lubin, Larsen, Matarazzo & Seever, 1986; Watkins, Campbell & McGregor, 1988). Because of this popularity, the Wechsler scales are taught in 88 percent of graduate level intelligence testing courses (Oakland & Zimmerman, 1986). Despite the prevalence of use of the Wechsler scales in practice and the classroom, there is a large body of evidence that shows examiner error is common. Errors are made by graduate students (Alfonso, Johnson & Patinella, 1998; Belk, 2001; LoBello & Holley,

1999; Slate, Jones & Murray, 1991), as well as experienced practitioners (Bradley, Hanna & Lucas, 1980; Whitten, Slate, Jones, Shine & Raggio, 1994). Researchers have identified different types of errors, distinguishing between administration, scoring, and clerical errors (Klassen & Kishor, 1996).

The large number of errors made by graduate students and practitioners illustrates the need to constantly re-evaluate and improve teaching methods. One aspect of instruction found in almost all methods of teaching individual intelligence testing courses is practice administrations. In a survey of 49 directors of psychology graduate programs, Oakland and Zimmerman (1986) found that programs required an average of 6.7 practice administrations of the Wechsler Intelligence Scale for Children-Revised Edition (WISC-R, Wechsler, 1974), 3.9 administrations of the Wechsler Adult Intelligence Scale-Revised Edition (WAIS-R, Wechsler, 1981) and 3.5 administrations of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-R, Wechsler, 1989). Regardless of the model of instruction, whether it is the traditional model, a competency-based model (Blakely, Fantuzzo & Moon, 1985; Fantuzzo, Sisemore & Spradlin, 1983), or clinical clerkship training (Moon et al., 1986), practice administrations are the norm.

Given the prevalence of practice administrations as a teaching method for the Wechsler scales of intelligence, researchers have sought to investigate their effectiveness (Conner & Woodall, 1983; Slate & Jones, 1989a; Slate et al., 1991; Slate et al., 1993). Most studies show little if any improvement with increased practice administrations on any of the Wechsler scales. Those studies that do show improvement do not find it consistently across different types of errors (Conner & Woodall, 1983).

The remainder of this chapter reviews the contemporary literature regarding the effects of practice administrations on the Wechsler scales of intelligence. A review of the types of errors made and their prevalence will be followed by an explanation of the primary causes of these errors. The ultimate goal of this study was to investigate the effects of practice administrations on students' ability to administer the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III). Unlike all other studies in this area, this study investigated the effects of practice administrations in the absence of structured feedback from the instructor. The review concludes with a discussion of the specific research questions examined in this study.

Prevalence of Use of The Wechsler Scales of Intelligence

There is no research showing the frequency of use of the WAIS-III, but many studies show that the WAIS-R, as well as all the Wechsler scales remain some of the most highly used tests of intelligence across all disciplines and settings (Archer et al., 1991; Lees-Haley, 1992; Lubin et al., 1986; Watkins et al., 1988). None of the previously mentioned studies isolated the use of adult intelligence tests from those that are given to children. Also, most of these studies did not measure the prevalence of intelligence tests specifically, but instead rated the use of many types of psychological tests. Kaufman (1990) sought to measure the frequency of intelligence test use by those who assess adults. 290 out of 300 (97 percent) surveys from respondents who test adults reported using the WAIS or WAIS-R. The second closest test was the Stanford Binet reported by only 25 percent. Camara, Nathan, & Puente (2000) performed the most recent study assessing the frequency of use of the Wechsler scales. Camara et al. found that clinical psychologists used the WAIS-R more than any other psychological test.

They also found that neuropsychologists used it second only to the MMPI-2. Harrison, Kaufman, Hickman and Kaufman (1988) found that most practitioners reported using the WAIS or WAIS-R primarily because both tests yield the most important information. They also considered the norms and theoretical soundness to be great strengths. Given the extreme past popularity of the WAIS and WAIS-R with users of intelligence tests, it is safe to assume that the WAIS-III shares the same high rate of use (Kaufman & Lightenberger, 2001)

Types of Examiner Errors

Administration Errors. Administration is the gateway to error. "You cannot interpret what you cannot score and cannot score what you cannot administer" (Moon et al., 1986). Administration errors as defined by Klasen and Kishor (1996) occur when the examiner fails to follow the standardized testing procedures. Examples of administration errors include not reading subtest instructions verbatim, failure to establish proper basal levels, failure to recognize points of discontinuation, and failure to time items when required. Failure to query the examinee for additional information when required or querying when not necessary is also a common administration error. Many of these errors can only be discovered during direct observations of testing (Stewart, 1987), but some, such as improper basal/ceiling levels and failure to query can be discovered upon review of protocols (Belk, 2001).

Scoring Errors. Scoring errors occur when an examiner fails to properly assign the correct point value for a subject's response (Klausen and Kishor, 1996). Numerous subtests on the WAIS-R and WAIS-III require subjective scoring and are therefore quite susceptible to errors. Vocabulary is one such subtest. Depending on what answer the

subject gives, the examiner must decide if the response earns one, two, or zero points based on the criteria provided in the test manual. Subtests like these must be scored while the test is being administered. Scoring "on-line" allows the examiner to discontinue the test when the proper criteria are met. This scoring method requires great expertise and familiarity with the scoring criteria. Other subtests, such as Arithmetic or Information usually have but one right answer. These tests are much easier to score.

Clerical Errors. Clerical errors are careless mistakes such as adding columns of numbers incorrectly, adding in optional subtests, or using incorrect norms tables (Belk, 2001). The most common error of this type regardless of classification is the clerical error of *failure to record* answers or response times. Incorrectly calculating chronological age is also a clerical error. There are a number of studies that show a prolific number of scoring and clerical errors committed by graduate students as well as practitioners in the administration of the WAIS-R (Alfonso, Johnson & Patinella, 1998; Belk, 2001; LoBello & Holley, 1999; Slate, Jones & Murray, 1991).

Frequency and Impact of Errors in Administering and Scoring the Wechsler Scales

Due to Wechsler scale's prevalence of use in education and psychology, there is a large body of evidence documenting errors made by graduate students while learning to administer these scales (Alfonso et al., 1998; Bradley et al., 1980; Slate & Hunnicutt, 1988; Slate et al., 1991) as well as practitioners (Bradley et al., 1980; Slate, Jones, Coulter & Covert, 1991; Whitten et al., 1994).

To identify examiner errors on the WAIS-R, Slate and Jones (1989b) reviewed 149 protocols submitted by 22 graduate students enrolled in an individual intelligence-testing course. Slate and Jones sought not only to record the frequency of errors (as

previous studies had done), but to identify and discriminate between different types of errors. Results showed that only four of the protocols were free of errors and that all 22 students had at least one protocol with errors. There was an average of 7.95 errors per protocol. The most common error made was the scoring error of awarding too much credit for a response. This error was made three times as often as the error of awarding too little credit for a response. This error shows a trend that is consistent among many different types of intelligence tests (Platt, 2001). The clerical error of not recording responses verbatim was the second most common error. Also prevalent were the errors of failing to query when necessary and querying when prohibited. The Vocabulary, Similarities, and Information subtests were the source of most errors. After corrections were made, it was found that the students overestimated 56 percent of the IQs by a range of 1-10 points. Underestimations of 1-2 points were made on 16 percent of the protocols.

Slate, et al., (1993) investigated errors made by practitioners on the WAIS-R by reviewing 50 protocols taken from psychological folders in a southern state. Eight different practitioners were represented, each of which had the minimum requirements of at least a masters degree, and internship, and supervised experience in intelligence testing. All 50 protocols were found to have errors. When *failure to record* errors were included, an astonishing 36.9 errors per protocol were found. *Failure to record* errors were found most often on the Digit Span, Picture Completion, and Vocabulary subtests. When *failure to record* errors were excluded, the mean number of errors per protocol was reduced to 15.4. After this exclusion, most errors were found on the Vocabulary, Comprehension, and Similarities subtests. After *failure to record* errors the next most common error was failure to query. Similar to the graduate students, practitioners were

more likely to fail to query when required than they were to query when prohibited. Also in common with studies assessing student errors was the practitioner's difficulty with assigning correct point values. IQ scores had to be corrected on 27 or 54 percent of the protocols. Eighty-eight percent of the IQs were corrected to be lower than the IQ determined by the practitioner. All of the changed IQs were within 5 points of the original IQ, however, the IQ of two clients that were originally above 70 were below 70 after corrections were made.

To investigate scoring reliability for practitioners and students, Ryan, Prifitera, & Powers (1983) had 19 practicing psychologists and 20 graduate students review the same two WAIS-R protocols. The practicing psychologists had an average of 7.3 years of experience and the graduate students all had at least a master's degree and had taken a course in intelligence testing. The protocols used were not confabulated; they were actual protocols from two vocational-counseling clients. Results showed tremendous variability in the examinee's IQs (4-18 points). Combining the performance on both protocols revealed that 35 percent of the student group was in perfect agreement with the actual Full Scale IQ and 32 percent of psychologists were in full agreement. It should be noted that perfect agreement with the actual IQ does not eliminate the possibility of scoring or clerical error. Scoring errors on one subtest could compensate for opposing errors on other subtests, therefore resulting in the correct IQ. These results dramatically show that scoring errors are made frequently by examiners regardless of experience and that they can have a profound effect on the accuracy of the results.

Studies on Wechsler scales other than the WAIS-R have shown similar trends in administration and scoring errors. Slate and Chick (1989) found an average of 15.2

errors per WISC-R protocol submitted by graduate students in a graduate level psychology program. Rates and types of errors were almost identical to the results found by Slate and Jones (1989b) when examining the WAIS-R. The most common type of error made was the error of assigning incorrect points to items. As with the WAIS-R, students assigned more points than necessary three times more than they assigned fewer points. Also similar to the WAIS-R study was the frequency of errors in querying responses and failure to record responses verbatim. Again, errors were most commonly found on the Vocabulary, Comprehension, and Similarities subtests.

Belk (2001) reviewed 100 WISC-III protocols submitted by graduate students. None of the 100 protocols were free of errors. Excluding *failure to record* errors, an average of 10.9 errors was found on each protocol. As with studies on other Wechsler scales, Belk found the most common errors to be failing to query when required, assigning too many points to a response, and assigning too few points to a response. The impact of all these errors meant that Full Scale IQs were overestimated on 46 percent of the protocols and underestimated on 21 percent. Full Scale IQs on seven protocols were changed by nine or more points. These changes lead to IQ classification changes on 11 protocols.

Similar to the WISC-R and WAIS-R, the WPPSI-R has also been studied in order to identify the rate and type of examiner errors made in administration. LoBello and Holley (1999) reviewed 121 WPPSI-R protocols submitted by 25 students enrolled in a graduate level individual intelligence-testing course. No protocols were found to be free of errors, although some contained only a few of the *failure to record* variety. Including *failure to record* errors resulted in a mean of 57.92 errors per protocol. Excluding *failure*

to record errors reduced the mean to 13.02 errors per protocol. The amount of scoring and clerical errors found was low, but the authors note that these errors can have a drastic impact on calculated IQ values (p. 18). Most errors committed were administration and scoring errors. Excluding *failure to record* errors, the results were similar to other studies performed on other Wechsler scales. The most common error made was assigning too many points to a response. 96 percent of the examiners made this error. The second most common error was not querying when required by the instructional manual, which was committed by 92 percent of the examiners. When examining total errors, the Arithmetic and Picture Completion had the highest number of errors. After eliminating *failure to record* errors, these subtests were found to have the least amount of errors implying that *failure to record* is the primary source of error on these subtests. After the exclusion of the *failure to record* errors, the subtests of Geometric Design, Vocabulary, and Mazes are found to be the most prone to errors. After making corrections, 75 percent of the Full Scale IQs were changed. Full scale IQs were overestimated on 48 percent of the protocols and underestimated on 27 percent. The average change in IQ regardless of direction was 2.4 points with a difference of five or more points on 16 protocols. The Full Scale IQ classification was changed on 18 of the protocols (LoBello & Holley, 1999).

Whitten et al. (1994) examined errors made by practitioners on the WPPSI-R and found similar results. Two qualified graduate assistants reviewed 57 WPPSI-R protocols obtained from a regional medical center and from a mental health center. The tests had been performed by doctoral level interns, post-master's level doctoral students, and doctoral level practitioners. The examiners committed errors on all 57 protocols.

Including *failure to record* errors, there was a mean of 73.2 errors per protocol. After adjusting the count for *failure to record* errors there was a mean of 27.1 errors per protocol. After *failure to record,* the next most common error was assigning too many points for a response. Assigning too many points to a response occurred 1.6 times more often than the error of assigning to few points. Other common errors included, failure to query when required to do so, which occurred 2.3 times more than the opposite error of querying when not necessary. When counting *failure to record* errors, Picture Completion, Information, and Mazes ranked as the three highest error prone subtests. Excluding *failure to record* errors changed the order to Mazes, Geometric Design, and Vocabulary. Corrections to errors caused changes in 31 of the protocols. Examiners were 1.4 times more likely to overestimate the Full Scale IQ than they were to underestimate it.

Causes of Examiner Errors

The bulk of the research on examiner error has focused on identifying shortfalls in training and suggesting and implementing possible alternatives (Fantuzzo et al., 1983; Moon et al., 1986; Slate et al., 1991; Stewart, 1987). Training for administering intelligence tests normally includes a practice demonstration and a discussion of administration and scoring procedures followed by repeated practice administrations with feedback from the instructor (Slate & Jones, 1990b). This "practice makes perfect" method has been shown to reduce some types of errors, but many errors remain, even after numerous administrations. If this were not true then we would expect to see more accurate administrations performed by experienced practitioners. This is clearly not the case (Bradley et al., 1980; Ryan et al., 1983; Slate et al., 1983).

Another major cause of examiner error as identified by Slate and Hunnicutt (1988) is ambiguity of the testing materials. Many of the Wechsler subtests are difficult to score within the guidelines given in the manuals. Numerous studies have shown that most errors occur on the Vocabulary, Comprehension, Similarities, and Information subtests (Blakely et al. 1985; Moon et al. 1986; Slate and Jones, 1990b). These are the tests that require the most subjective scoring by the examiner and are the source of most of the errors involving incorrect point assignment and querying. Researchers from Miller and Chansky (1972) to Whitten et al. (1994) have recommended changes in the Wechsler manuals that would simplify scoring as well as supplemental response listings.

Examiner carelessness is also a large source of error. This carelessness most commonly leads to scoring and clerical errors, but can also be a major source of administration error. Jones and Slate note that examiner carelessness has been observed particularly in practitioners who are under time constraints to complete numerous tests (as cited in Whitten et al., 1994).

The Effects of Practice in the Administration of the Wechsler Scales

"Practice makes perfect" has been the mantra for almost all courses that specialize in teaching the administration of the Wechsler scales. Because of this, many studies have been conducted to assess the effects of practice administrations on the number of errors made by examiners. For the most part, there have been two strategies used in investigating the effects of practice administrations on examiner errors. Some studies have measured the performance of examiners on one Wechsler scale after practicing with another (Platt, 2002; Slate et al., 1991, Slate & Jones, 1990b). Others simply measure the prevalence of errors after multiple practice administrations of the same scale (Alfonso et

al., 1998; Conner & Woodall, 1983). The rationale for the former method comes from the idea that there is a considerable amount of positive transfer from one Wechsler scale to another (Slate et al., 1991).

Conner and Woodall (1983) studied the effects of practice administrations and structured feedback using the WISC-R. Ten graduate students administered and scored 15 tests. The student's protocols were organized into groups of five. The first group contained the first through fifth administration, the second group contained the sixth through tenth, and the third group consisted of the eleventh through fifteenth. For the purposes of their study, Conner and Woodall categorized errors as Response scoring, IQ, Administrative, and Mathematical. With experience, the total number of errors, as well as the number of administration errors, decreased significantly. The mean number of total errors in group three was half of that in group one. The mean number of administration errors in group one was 2.6, compared to a group three mean of .90. However, other categories of errors such as mathematical, scoring, and IQ remained constant.

Examining the Wechsler Intelligence Scale for Children-Third Edition (WISC-III, Wechsler, 1991), Alfonso et al. (1998) reviewed 60 protocols administered by graduate students to measure the frequency and types of errors made. Each student was required to administer the WISC-III four times. Prior to administration, the students were given classroom instruction on the test stimuli, the different subtests, and a review of the most common errors made when administering the Wechsler scales. A thorough review of the WISC-III protocol was also performed. Students were given extensive verbal and written feedback between each of the four test administrations. Errors were found on all the

protocols. There was an average of 7.8 errors per protocol. The most notable aspect of this study is the average decrease in errors from the first administration (14.4) to the fourth (5.4). Also, when adding together the total number of errors made in the five most common categories of failure to query, *failure to record*, reporting FSIQ incorrectly, reporting VIQ incorrectly, and adding individual subtest scored incorrectly, the total number of errors dropped from 76 on the first administration to 27 on the fourth. These results call into question the common practice of many intelligence testing course instructors that requires students to administer five or more tests (p. 124).

Slate and Jones (1990b) analyzed 180 WAIS-R protocols from 26 graduate student examiners for administration, clerical, and scoring errors. Fifteen students administered the WAIS-R eight times and then administered the WISC-R five times. The remaining 11 students administered the WAIS-R five times before administering the WISC-R eight times. Verbal and written feedback was given to each student between administrations by a qualified graduate student who had already finished the course. Students averaged 8.8 errors per protocol but only three of the 180 protocols were error free. This time, the analysis showed that the most common type of error made was failing to record responses verbatim. This error however is not as detrimental to the correct scoring of the test as it is to later clinical analysis of the results. The next most frequent error was the scoring error of incorrect point assignment. Again, students were three times more likely to assign too many points than they were to assign too few. Slate and Jones noted that this resulted in numerous incorrect subtest raw scores and inflated IQs. The administration error of inappropriate querying was also found in many of the protocols. 160 of the protocols reviewed had at least one instance of the examiner failing

to query when necessary, and 106 had instances of unnecessary querying. In this study, 81 percent of the IQs were changed after errors were corrected. Sixty-two percent of the students overestimated the IQs while 17.8 percent underestimated it. Sixty-four percent of the protocols had an IQ difference of over 2 points and 16 percent had a difference of 3 points or more. This study also shows how administration, scoring, and clerical errors can add up to have a tremendously adverse impact on the reliability of the determined IQ. No improvement was found after five practice administrations, however, significant improvement was found after eight. Even with this improvement, many errors were still committed. The students in this study appeared to be practicing bad testing habits rather than improving their skills (p. 86).

Slate et al. performed a similar study in 1991. They analyzed 150 WAIS-R protocols submitted by 20 graduate students to examine the effects of practice administrations in teaching the WAIS-R. One half of the students were assigned to administer the WISC-R five times followed by 10 WAIS-R administrations. The other half of the students administered the WISC-R ten times before administering the WAIS-R 5 times. The students received oral feedback after their practice administrations. An examination of the amount of errors made on the first WAIS-R administration was used to assess the effectiveness of practice administrations with the WISC-R. No significant differences were found between the group that gave 10 WISC-R practice administrations and the group that gave 5 practice administrations. This was true with or without the inclusion of *failure to record* errors. The authors also investigated the number of errors made across test administrations. Including *failure to record* errors there was a significant decrease in errors across 10 test administrations, but not across five

administrations. Excluding *failure to record* errors found no decrease across ten practice administrations and an increase in errors across five. When the overall error totals were adjusted to remove recording errors, most errors were found on the Vocabulary, Similarities, Picture Completion, and Information subtests. This is consistent with almost all other studies on the Wechsler scales. Also consistent with other studies was the observation that the errors of assigning too many or too few points and incorrect querying procedures were most common. After corrections were made, only 12 percent of the Full Scale IQs calculated by the students were unchanged. The students overestimated 84 percent of the Full Scale IQs and underestimated 4 percent.

Given the obvious inconsistency of the effectiveness of practice administrations in the reduction of examiner errors on the Wechsler scales of intelligence, some researchers have sought to compare teaching methods and look for improvements (Slate and Jones, 1990a; Slate, Jones & Covert, 1992). One common aspect of all the previously mentioned studies investigating the effects of practice administrations on the rate of examiner errors is feedback. This variable makes it rather difficult to assess the true effects of practice administrations since feedback can be expected to partially explain some of the improvements across numerous administrations. Feedback is but one of the many aspects of the differing teaching methods that can have an effect on examiner errors. Some instructional methods have taken a more objective approach and rely on the use of checklists (Blakely et al., 1985; Fantuzzo et al., 1983; Fantuzzo & Moon, 1984). Other studies have been performed to investigate the effectiveness of automated models of instruction (Blakely et al., 1985) and peer mediated instruction (Blakely, Fantuzzo, Gorsuch, and Moon, 1987).

This review of the literature shows that the effects of practice administrations as a teaching method are rather inconsistent. Some studies have shown that administrative errors can be reduced, but scoring and clerical errors remain constant. Other studies even imply that practice administrations cause the examiner to practice and reinforce bad testing procedures instead of improve and remove errors.

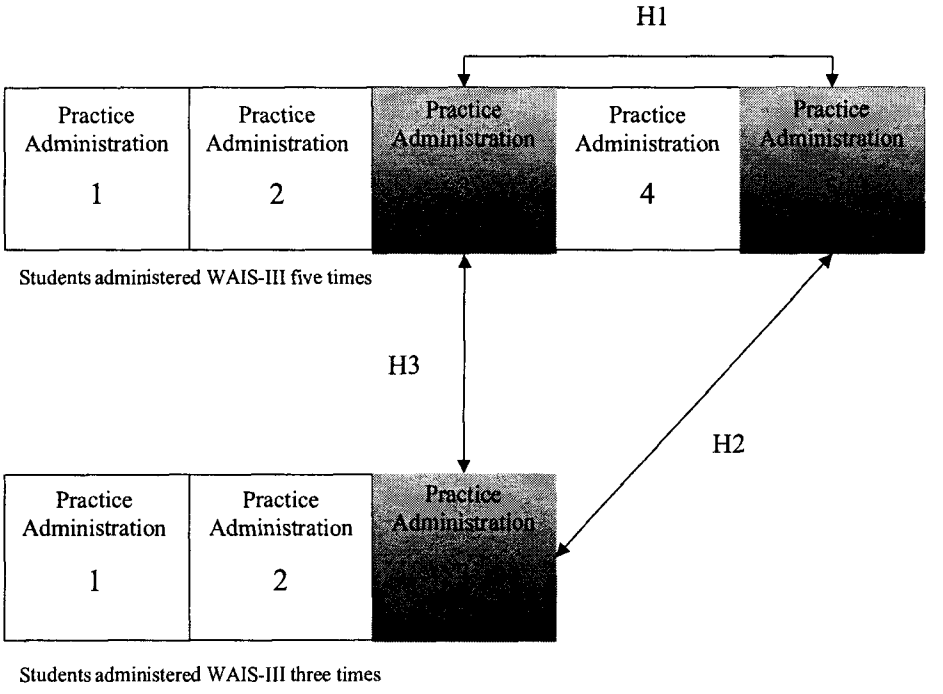
This study investigated the effects of practice administrations on the rate of total errors and adjusted errors made by student examiners on the WAIS-III in the absence of structured feedback. Adjusted errors were defined as the amount of errors excluding *failure to record* errors. Specifically, I examined the differences in error rates between three and five practice administrations of the WAIS-III. One goal of this study was to find the effects of practice administrations in the absence of feedback. Another was to attempt to identify a minimum number of practice administrations needed to see a decrease in examiner error. Administration, scoring and clerical errors were evaluated on a sample of WAIS-III protocols submitted by graduate students enrolled in an individual intelligence testing course. The author hypothesized that:

H1. Without structured feedback between administrations, students required to perform 5 WAIS-III administrations would commit fewer total and adjusted errors on their fifth administration than on their third.

H2. Without structured feedback between administrations, students required to perform 5 WAIS-III administrations would commit fewer total and adjusted errors on their last administration than the last administration of another group required to perform 3 administrations.

H3. Without structured feedback between administrations, there would be no difference in the average number of total and adjusted errors committed on the third administration of the WAIS-III between groups (See Figure 1).

Figure 1 - Hypotheses



Method

Participants

14 graduate students enrolled in 4 sections of a graduate course in intelligence testing submitted a total of 56 WAIS-III protocols for evaluation. The students were in the first year of a master's degree program in applied psychology. None had previous experience in individual intelligence testing, but all had completed a basic course in psychometrics. Ten of the students were female and four were male. Of the ten female students, eight were White, and two were black. Three of the male students were White and one was East Indian.

The Intelligence testing course was structured so that the students learned to administer either the WAIS-III or the WISC-III first. For the purposes of this study, only protocols from those students that administered the WAIS-III first were used. Of the 14 students, seven administered the WAIS-III three times (21 protocols) and seven administered the WAIS-III five times (35 protocols). Each student in each section of the class turned in all the WAIS-III protocols at the same time. No formal feedback was given between test administrations.

Of the seven students that administered the WAIS-III five times, only three administered all the tests on separate days. Three students gave two tests on the same day and one student administered three in one day. In the group that administered the WAIS-III three times, four students gave all the tests on separate days. The other three students each administered two tests on the same day. It is impossible to determine the order of testing in these cases. To overcome this difficulty, each day was considered a testing session. Tests given on the same day were assigned to the same group for

analysis. The mean of the tests given during the same day or session was used for the computations. The rationale is that giving multiple tests in one day would not allow for review of the manual and testing procedures and therefore would not lead to improved accuracy. A major disadvantage of this procedure is that it truncates the higher end of the distribution of protocols so that there is only three protocols in the fifth administration condition. Figure 2 shows the remaining number of protocols to be used in each condition.

Figure 2 - Number of Protocols

Practice Administration 1	Practice Administration 2	Practice Administration 3	Practice Administration 4	Practice Administration 5
n = 7	n = 7	n = 7	n = 6	n = 3

Students administered WAIS-III five times

Practice Administration 1	Practice Administration 2	Practice Administration 3
n = 7	n = 7	n = 4

Students administered WAIS-III three times

Materials

A checklist was created using the requirements for correct administration and scoring from the WAIS-III manual (Wechsler, 1997) was used to evaluate the errors made by the student examiners. There is a separate checklist for each subtest except for

the optional Object Assembly subtest (see Appendix A). A separate checklist was used to check chronological age calculations and to record any corrections made to subtest scores or IQs. Also, the optional subtests of Object Assembly, Digit Symbol-Incidental Learning and Free Recall were not included in this study.

Procedure

All students received lectures on the proper administration and scoring of the Wechsler scales of intelligence prior to beginning administrations. Each student was instructed to read and become familiar with the instruction manual before testing. Two graduate assistants who had previously completed the same intelligence-testing course reviewed and re-scored the protocols using the checklists. Twenty of the protocols were reviewed by one student and the remaining 36 were reviewed by another.

In order to obtain inter-scorer reliability, after all protocols had been re-scored, five were randomly selected and reviewed a third time by another graduate assistant. The third graduate assistant has also previously completed the intelligence testing course and has worked for one year in a local outpatient mental health facility in the position of psychometrist. When any disagreement occurred between the scorers that changed a score on a test, the disagreement was reconciled and all tests were re-scored to reflect the reconciliation. This process was repeated until there was no disagreement that altered any I.Q. score, index score, or subtest scaled score in a set of five tests. When this criterion was met, the reliability check was discontinued. This criterion was met after ten protocols were jointly evaluated.

A conservative criteria for identifying errors was used. Only responses and administration procedures that specifically violate the guidelines exhibited in the WAIS-

III manual were considered errors. Illegibly recorded responses were not counted as errors. Finally, when the examiner did not write down the verbatim response given by the examinee, record the response time, or makes similar omissions, these errors were classified as *failure to record* errors (Belk, 2001).

Results

Errors were committed on all of the 56 WAIS-III protocols. With failure to record errors included, students who administered the WAIS-III only three times averaged 31.4 errors per protocol (range = 4.0 to 82.0). When *failure to record* errors were excluded, the number of errors was reduced to 10 per protocol (range = 3.0 to 18.0). Including *failure to record* errors, students who administered the WAIS-III five times committed an average of 23.0 errors per protocol (range 4.0 to 82.0). Excluding failure to record errors yielded 15.0 errors per protocol (range 4.0 to 20.0).

One-way analyses of variance were completed to evaluate significant changes in errors made by students across test administrations. Table 1 includes all the means for adjusted and total errors for both groups and will be referenced for each hypothesis test. Including *failure to record* errors, no significant differences were found across administrations in the five administration group or the three administration group. The same is true when excluding *failure to record* errors.

The first hypothesis that, in the absence of structured feedback, students required to perform 5 WAIS-III administrations would commit fewer total and adjusted errors on their fifth administration than on their third was not confirmed. Using a one-way analysis of variance with a critical value of .05, significant differences were not found for unadjusted errors ($F_{4, 30} = .77$, $p = .55$) or adjusted errors ($F_{4, 30} = .88$, $p = .489$). Therefore, the amount of errors did not decrease between three and five administrations.

The second hypothesis stated that, in the absence of structured feedback, students required to perform 5 WAIS-III administrations would commit fewer total and adjusted errors on their fifth and final administration than the third and final administration of

Table 1

Error Means for 5 and 3 Practice Administrations

Errors/# of Admins	*Test 1*	*Test 2*	*Test 3*	*Test 4*	*Test 5*
Unadjusted Errors					
5	n=7 **30.6** (27.6)	=11 **29.4** (21.1)	n=8 **24.3** (14.6)	n=6 **18.3** (12.0)	n=3 **12.3** (11.0)
3	n=8 **31.8** (24.9)	n=9 **39.6** (30.5)	n=4 **9.5** (4.2)		
Adjusted Errors					
5	n=7 **17.6** (6.1)	n=11 **15.5** (5.8)	n=8 **17.5** (7.9)	n=6 **14.5** (6.2)	n=3 **10.0** (8.7)
3	n=8 **12.1** (5.1)	n=9 **9.3** (2.8)	n=4 **9.5** (4.2)		

Note. Standard deviations are in parentheses.

[1] *Failure to record* errors included. [2] *Failure to record* errors excluded.

another group required to perform only 3 test administrations. This hypothesis was also not supported. An independent sample t-test was performed which showed that there was no difference in the amount of unadjusted errors ($t_5 = .72$, $p = .5$) errors or adjusted errors ($t_5 = -.10$, $p = .92$) committed after performing 5 practice administrations and performing 3 practice administrations. Table 1 shows the mean number of errors committed in each condition.

The third hypothesis stated that, without structured feedback, there will be no difference in the average amount of total and adjusted errors committed on the third administration of the WAIS-III between groups was supported. This was used as a between groups reliability check. If all other variables are equal you would not expect to see a difference when comparing these two conditions. Another independent sample t-test was performed which showed that there was no difference in the number of unadjusted errors ($t_{10} = -.12$, $p = .91$) errors or adjusted errors ($t_{10} = -1.8$, $p = .09$) between the third administrations of each group.

Discussion

The overarching hypothesis of this study was that multiple practice administrations of the WAIS-III by student examiners would reduce the number of total and adjusted errors committed by those students. Two additional administrations did not cause a significant reduction in examiner errors. This can be seen in the results that errors did not decrease between the third and fifth administration for those students that administered five tests. There also was no difference between the last (fifth) administration of the group that administered five tests and the last (third) administration of the group that administered three tests. This study provided no evidence to show that practice administrations without structured feedback reduce errors. These results are consistent with those found by Slate et al. (1991). In their study, they found no improvement over five administrations of the WAIS-R. In fact, they reported an increase in errors in the fifth administration, when compared to the first. Over ten administrations, Slate found a decrease in total errors that he attributed solely to the reduction in *failure to record* errors. However even after ten administrations, the students in Slate et al's study made an average of 13 errors per protocol.

The assumption for years in intelligence testing classes has been that "practice makes perfect". It is believed that increased numbers of practice administrations will decrease the number of errors made by student examiners. This assumption appears to be incorrect. No improvement was seen between three and five practice administrations. In fact, no significant decrease in total or adjusted errors was seen in either group across any of the three or five administrations. It may be true that increasing the number of administrations may only serve to ingrain bad habits learned on early administrations.

Previous studies that have shown mixed results or positive results for practice administrations have all used some form of structured feedback between administrations. Because this study did not use structured feedback and did not show significant decreases in errors, the importance of feedback in the training of student examiners should be considered. Slate et al. (1991) recommend that his feedback be immediate and specific. The teaching model used in this study used organized feedback only after students had administered all five WAIS-III or WISC-III protocols. Practice administrations were relied on heavily as a method of teaching, however, much time was spent in class (which met twice a week for 1.25 hrs) exposing the students to those areas most susceptible to administration and scoring errors. The instructor also spent time addressing specific subtests and subtest items that are known to be the source of may errors. Students were administering the WAIS-III concurrently with these class meetings over a period of approximately two weeks.

Another possible explanation for not finding a significant reduction in total and adjusted errors across practice administrations may lie in the small sample sizes in some of the higher end conditions. The truncation of the distribution due to tests being administered by participants on the same day caused reduced sample sizes in these conditions. If possible, more protocols should be used to increase sample size, in order to obtain an increase in statistical power. In other words, we don't really know if there is "no significant difference" or if the sample sizes were so small that statistical tests were incapable of detecting differences.

There were other limitations to this design as in addition to the relatively small sample size. The protocols were obtained from previous intelligence testing classes and

were used because of their convenience for the study. Because all of the possible protocols were used random assignment of protocols to the various conditions was not possible.

There may also be some inconsistencies in how the four sections of the intelligence-testing course was taught. They were all taught by the same instructor, however, because in two of the sections students were required to administer five tests and the other two sections were only required to administer three there could be some small inconsistencies in how the material was presented. Due to the previously mentioned limitations in controlling and manipulating all the relevant variables this study should be considered as having a quasi-experimental design. The threats to internal validity that need to be discussed in this study are history and selection. Because the protocols used were from four different sections of the individual intelligence-testing class it is impossible to know what specific events may have happened to some of the student examiners and not to others. For instance, some students may have made special arrangements to meet with the instructor to ask questions. This would be considered a form of formal feedback and could have a profound impact on performance in administering the test. Also, because all the protocols available were used, selection bias may exist.

Ideally, this study would be performed using just one (or at the most two) sections of the individual intelligence-testing course. One half of the course would be assigned to perform five administrations of the WAIS-III and the other half would be assigned three administrations. This would make it possible to ensure that all the students were receiving the same amount and quality of instruction before and during administrations.

Each of the students would also be instructed to administer no more than one test per day to make sure that none of the protocols had to be eliminated as in this current study. Ideally, there would be enough students in the class to be able to randomly choose the students whose protocols would be used in the study.

In summary, the present study offered no evidence that, in the absence of structured feedback, increased practice administrations of the WAIS-III by student examiners will reduce the number of either total or adjusted errors committed by those students. The results reinforce the importance of structured, immediate, and specific feedback as a method of teaching administration of the Wechsler scales of intelligence.

References

Alfonso, V. C., Johnson, A., & Patinella, L. (1998). Common WISC-III examiner errors: Evidence from graduate students in training. Psychology in the Schools, 35 (2), 119-125.

Archer, R. P., Maruish, M., Imhof, E. A., & Piotrowski, C. (1991). Psychological test usage with adolescent clients: 1990 survey findings. Professional Psychology: Research and Practice, 22 (3), 247-252.

Belk, M. S. (2001). WISC-III adminstration, clerical, and scoring errors made by student examiners. Unpublished master's thesis, Auburn University Montgomery.

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 Consulting and Clinical Psychology, 48, 530-531.

Camara, W.J., Nathan, J. S., Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. Professional Psychology: Research and Practice, 31, 141-154.

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. (1984). Mastery: A competency-based training model for clinical psychologists. The Clinical Psychologist, 37 (2), 29-30.

Fantuzzo, J. W., Sisemore, 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.

Hanna, G., Bradley, F., & Holen, M. (1981). Estimating major sources of measurement error in individual intelligence scales: Taking our heads out of the sand. Journal of School Psychology, 19 (4), 371-376.

Harrison, P.L., Kaufman, A. S., Hickman, J. A., & Kaufman, N. L. (1988). A survey of tests used for adult assessment. Journal of Psychoeducational Assessment, 6, 188-198

Hunnicutt, Jr., L., Slate, J., Gamble, C., & Wheeler, M. (1990). Examiner errors on the Kaufman Assessment Battery for Children: A preliminary investigation. Journal of School Psychology, 28, 271-278.

Kaufman, A. S. (1990). Assessing adolescent and adult intelligence. Boston, MA: Allyn & Bacon.

Kaufman, A. S., Lichtenberger, E. O. (2002). Assessing adolescent and adult intelligence 2nd ed. Boston, MA: Allyn & Bacon.

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

Lees-Haley, P. R. (1992). Psychodiagnostic Test Usage by Forensic Psychologists. American Journal of Forensic Psychology, 10 (1), 25-29.

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

Lubin, B., Larsen, R. M., Matarazzo, J. D., & Seever, M. (1986). Psychological test usage patterns in five professional settings. American Psychologist, 40, 857-861.

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

Moon, G. W., Fantuzzo, J. W., & Gorsuch, R. L. (1986). Teaching WAIS-R Administration Skills: Comparison of the Mastery Model to Other Existing Clinical Training Modalities. Professional Psychology: Research and Practice, 17, 31-35.

Oakland, T. D., & Zimmerman, S. A. (1986). The course on individual mental assessment: A national survey of course instructors. Professional School Psychology, 1 (1), 51-59.

Platt, T. L., (2002). The effects of practice and organized feedback on scoring and administration on Third Edition Wechsler scales. Unpublished master's thesis, Auburn University Montgomery.

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

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

Slate, J., & Hunnicutt, Jr., L. (1988). Examiner errors on the Wechsler Scales. Journal of Psychoeducational Assesment, 6, 280-288.

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

Slate, J., & Jones, C. (1989b). Examiner error on the WAIS-R: A source for concern. The Journal of Psychology, 124 (3), 343-345.

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

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

Slate, J., Jones, C., Coulter, C., & Covert, T. (1991). Practitioner's administration and scoring of the WISC-R: Evidence that we do err. Journal of School Psychology, 30 (1), 77-83

Slate, J., Jones, C., & Covert, T. (1992). Rethinking the instructional design for teaching the WISC-R: The effects of practice administration. College Student Journal, 26 (3), 285-289.

Slate, J., Jones, C., & Murray, R. (1991). Teaching administration and scoring of the Wechsler Adult Intelligence Scale-Revised: An empirical evaluation of practice administration. Professional Psychology: Research and Practice, 20, 375-379.

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

Stewart, K. J. (1987). Assessment of technical aspects of WISC-R administration. Psychology in the Schools, 24, 221-228.

Watkins, C. E., Campbell, V. L., & McGregor, P. (1988). Counseling psychologists' use of and opinions about psychological tests: A contemporary perspective. The Counseling Psychologist, 16 (3), 476-486.

Wechsler, D. (1974). Wechsler Intelligence Scale for Children-Revised. New York: The Psychological Corporation.

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

Wechsler, D. (1989). Wechsler Preschool and Primary Scale of Intelligence-Revised. San Antonio, TX: The Psychological Corporation.

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

Wechsler, D. (1997). Wechsler Adult Intelligence Scale-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 **WAIS-III** Protocols

Name of Student: _____ Date of test: ___/___/___.
Client birth date: ___/___/___.

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. _____ Sums of subtest scaled scores copied incorrectly from SCORE CONVERSION PAGE to PROFILE PAGE.

6. Addition errors:

VIQ	YES / NO	VCI	YES / NO
PIQ	YES / NO	POI	YES / NO
FSIQ	YES / NO	WMI	YES / NO
		PSI	YES/ NO

7. Wrong norms tables used: YES / NO

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

9. PIQ assigned by student _____
   Recalculated PIQ _____
   Difference (+/-) _____

10. VIQ assigned by student _____
    Recalculated Verbal IQ _____
    Difference (+/-) _____

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

12. Student VIQ-PIQ Difference _____
    Actual Difference _____

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

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

15. WMI assigned by student _____
    Recalculated WMI _____
    Difference (+/-) _____

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

1. PICTURE COMPLETION

	No. resp. recorded	0 pt. resp. given 1 pt.	1 pt. resp. given 0 pts.
1. Comb			
2. Table			
3. Face			
4. Briefcase			
5. Train			
6. Door			
7. Glasses			
8. Pitcher			
9. Pliers			
10. Leaf			
11. Pie			
12. Jogging			
13. Fireplace			
14. Mirror			
15. Chair			
16. Roses			
17. Knife			
18. Boat			
19. Basket			
20. Clothing			
21. Lockers			
22. Cow			
23. Tennis			
24. Woman			
25. Barn			

Starting point correct (Item #6): YES / NO
Basal level correct (perfect scores on 2 consecutive items): YES / NO
Ceiling level correct (5 consecutive scores of 0): YES / NO
Sum of item scores correct: YES / NO
Gave credit for items below basal level: YES / NO
Reversed sequence until examinee earned perfect scores (2-points) on 2 consecutive items (should be done if examinee earns scores of 0 on either item 6 or 7): YES / NO / N.A.

2. VOCABULARY

	Should have (Q)'d	(Q)'d when not needed	Gave 2 pts. for 1 pt. resp.	Gave 1 pt. for 2 pt. resp.	Gave any pts. for 0 pt. resp.	No resp. recorded	1 or 2 pt. resp. scored 0
1. Bed							
2. Ship							
3. Penny							
4. Winter							
5. Breakfast							
6. Repair							
7. Assemble							
8. Yesterday							
9. Terminate							
10. Consume							
11. Sentence							
12. Confide							
13. Remorse							
14. Ponder							
15. Compassion							
16. Tranquil							
17. Sanctuary							
18. Designate							
19. Reluctant							
20. Colony							
21. Generate							
22. Ballad							
23. Pout							
24. Plagiarize							
25. Diverse							
26. Evolve							
27. Tangible							
28. Fortitude							
29. Epic							
30. Audacious							
31. Ominous							
32. Encumber							
33. Tirade							

Starting point correct (item 4): YES / NO
Basal level correct (perfect scores on 2 consecutive items): YES / NO
Ceiling level correct (six consecutive scores of 0): YES / NO
Sum of item scores correct: YES / NO
Gave credit for items below basal level: YES / NO
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 item 4 or 5): YES / NO / N.A.

3. DIGIT SYMBOL-CODING

Completion time recorded:	YES / NO
Completion time 120" or less:	YES / NO
Total Raw Score recorded:	YES / NO
Total Raw Score correct:	YES / NO
Subject completed at least 4 rows during the allotted time:	YES / NO
If subject completed less than 4 rows during the allotted time, examiner marked point on protocol and allowed subject to continue:	YES / NO / N.A.
Used black lead pencil:	YES / NO

Number of scoring errors: _____

****

Digit Symbol-Incidental Learning

Pairing

Total Score Correct (Max=18):	YES / NO
Used black lead pencil:	YES / NO

Free Recall

Total Score Correct (Max=9):	YES / NO
Used black lead pencil:	YES / NO

****

Digit Symbol-Copy

Completion Time Recorded:	YES / NO
Completion Time 90" or less:	YES / NO
Total Raw Score Recorded:	YES / NO
Total Raw Score Correct:	YES / NO

Number of Scoring Errors: _____

Errors Due to Error on Scoring Key: YES / NO

4. SIMILARITIES

	Should have (Q)'d	(Q)'d when not needed	Gave 2 pts. for 1 pt. resp.	Gave 1 pt. for 2 pt. resp.	Gave any pts. for 0 pt. resp.	Gave 1 or 2 pt. resp. 0 pts.	No resp. recorded
1. Fork-Spoon							
2. Socks-Shoe							
3. Yellow-Green							
4. Dog-Lion							
5. Coat-Suit							
6. Piano-Drum							
7. Orange-Banana							
8. Eye-Ear							
9. Boat-Automobile							
10. Table-Chair							
11. Work-Play							
12. Steam-Fog							
13. Egg-Seed							
14. Demo-Monarch							
15. Poem-Statue							
16. Praise-Punish							
17. Fly-Tree							
18. Hiber-Migra							
19. Enemy-Friend							

Starting point correct (item 6): YES / NO
Basal level correct (perfect scores on 2 consecutive items): YES / NO
Ceiling level correct (4 consecutive scores of 0): YES / NO
Sum of item scores correct: YES/ NO
Gave credit for items below basal level: YES NO N.A.
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 items 6 or 7): YES / NO / N.A.

5. BLOCK DESIGN

		Time not recorded	Item score not circled	2nd trial not given when required	2nd trial given when not required	Y / N not circled	Incorrect score circled	Bonus pts. incorrect
1	Trial 1							
	Trial 2							
2	Trial 1							
	Trial 2							
3	Trial 1							
	Trial 2							
4	Trial 1							
	Trial 2							
5	Trial 1							
	Trial 2							
6	Trial 1							
	Trial 2							
7								
8								
9								
10								
11								
12								
13								
14								

Starting point correct (item 5): YES / NO
Basal level correct (perfect scores on 2 consecutive items): YES / NO
Ceiling level correct (3 consecutive scores of 0): YES / NO
Sum of Item Scores correct: YES / NO
Gave credit for items below basal level: YES / NO
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 items 5 or 6): YES / NO / N.A.

6. ARITHMETIC

	No resp. recorded	0 pts. given for 1 pt. resp.	1 pt. given for 0 pt. resp.	Time not recorded	Bonus pts. not awarded
1. 3					
2. 7					
3. 5					
4. 2					
5. $9.00					
6. $4.00					
7. 5					
8. $1.50					
9. 8					
10. $3.60					
11. $10.50					
12. 30 cents					
13. $186.00					
14. 10					
15. $600.00					
16. 43					
17. $51.00					
18. $49.50					
19. 1 of 4; 5 of 20					
20. 96					

Starting point correct (item 5): YES / NO
Basal level correct (perfect scores on 2 consecutive items): YES / NO
Ceiling level correct (4 consecutive scores of 0): YES / NO
Sum of Item Scores correct: YES / NO
Gave credit for items below basal level: YES / NO
Reversed sequence until examinee earned perfect scores (2 points) on 2 consecutive items (should be done if examinee earns scores of 0 on items 5 and 6): YES / NO / N.A.

7. MATRIX REASONING

	Did not circle subject's resp.	Incorrect item score
A.		
B.		
C.		
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		
11.		
12.		
13.		
14.		
15.		
16.		
17.		
18.		
19.		
20.		
21.		
22.		
23.		
24.		
25.		
26.		

Administered sample items A, B, and C: YES / NO
Starting point correct (item 4): YES / NO
Basal level correct (perfect scores on 2 consecutive items): YES / NO
Ceiling level correct (4 consecutive scores of 0 or 4 scores of 0 on 5 consecutive items): YES / NO
Sum of Item Scores correct: YES / NO
Gave credit for items below basal level: YES / NO
Reversed sequence until examinee earned perfect scores (1 point) on 2 consecutive items (should be done if examinee earns scores of 0 on items 4 or 5): YES / NO / N.A.

8. DIGIT SPAN

Digits Forward

		No resp. recorded	Trial score not recorded	Item score not recorded	Incorrect trial score	Incorrect item score
1	1-7					
	6-3					
2	5-8-2					
	6-9-4					
3	6-4-3-9					
	7-2-8-6					
4	4-2-7-3-1					
	7-5-8-3-6					
5	6-1-9-4-7-3					
	3-9-2-4-8-7					
6	5-9-1-7-4-2-8					
	4-1-7-9-3-8-6					
7	5-8-1-9-2-6-4-7					
	3-8-2-9-5-1-7-4					
8	2-7-5-8-6-2-5-8-4					
	7-1-3-9-4-2-5-6-8					

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

Digits Backward

		No resp. recorded	Trial score not recorded	Item score not recorded	Trial score incorrect	Item score incorrect
1.	2-4					
	5-7					
2.	6-2-9					
	4-1-5					
3.	3-2-7-9					
	4-9-6-8					
4.	1-5-2-8-6					
	6-1-8-4-3					
5.	5-3-9-4-1-8					
	7-2-4-8-5-6					
6.	8-1-2-9-3-6-5					
	4-7-3-9-1-2-8					
7.	9-4-3-7-6-2-5-8					
	7-2-8-1-9-6-5-3					

Starting point correct (Item 1): YES / NO
Ceiling level correct (score of 0 on both trials of any item): YES / NO
Sum of Digits Backward item scores correct: YES / NO

Sum of Digits Forward + Digits Backward correct (Max= 30): YES / NO

9. INFORMATION

	No resp. recorded	1 pt. resp. given 0 pts.	0 pt. resp. given 1 point	Should have (Q)'d	(Q)'d when not needed
1. Saturday					
2. Age					
3. Ball					
4. Months					
5. Themo					
6. Sunrise					
7. Weeks					
8. Hamlet					
9. Brazil					
10. MLK					
11. Civ War					
12. Cleopatra					
13. Italy					
14. Relativity					
15. Olympics					
16. Sahara					
17. Genesis					
18. Sistine					
19. Gandhi					
20. Koran					
21. Water					
22. Vessels					
23. Cath					
24. Continent					
25. Curie					
26. World Pop					
27. Speed					
28. Faust					

Starting point correct (item 5): YES / NO
Basal level correct (perfect scores on 2 consecutive items): YES / NO
Ceiling level correct (6 consecutive scores of 0): YES / NO
Sum of item scores correct: YES/ NO
Gave credit for items below basal level: YES NO N.A.
Reversed sequence until examinee earned perfect scores (1 point) on 2 consecutive items (should be done if examinee earns scores of 0 on items 5 or 6): YES / NO / N.A.

10. PICTURE ARRANGEMENT

	Time not recorded	Resp. order not recorded	Score not circled	Incorrect score circled
1A. CAP				
1B. CAP				
2. BAKE				
3. OPENS				
4. CHASE				
5. CLEAN				
NCLEA				
6. HUNT				
THUN				
7. SAMUEL AMUELS				
SALMUE				
8. LUNCH				
LUCNH				
9. CHOIR				
HCOIR				
10. DREAM				
11. SHARK				

Starting point correct (Item 1): YES / NO
Ceiling level correct (four consecutive scores of 0 beginning with item 2): YES / NO / N/A
Sum of item scores correct: YES / NO

11. COMPREHENSION

	Should have (Q)'d	(Q)'d when not needed	Gave 2 pts. for 1 pt. resp.	Gave 1 pt. for 2 pt. resp.	Gave any pts. for 0 pt. resp.	No resp. recorded	1 or 2 pt. resp. given 0 points
1. Use Money							
2. Wear Watches							
3. Wash Clothes							
4. Envelope							
5. Food Cooked*							
6. Parole System *							
7. Child Labor*							
8. Professional Service							
9. Taxes							
10. History*							
11. Deaf							
12. Forest							
13. Jury Peers*							
14. City Land							
15. Marriage Lic							
16. Free Press							
17. Swallow							
18. Shallow Brooks							

*Requires two 1-point responses.
Did examiner (Q) for additional response to:

ITEM #5 YES / NO / N/A
ITEM #6 YES / NO / N/A
ITEM #7 YES / NO / N/A
ITEM # 10 YES / NO / N/A
ITEM #13 YES / NO / N/A

Starting point correct (item 4): YES / NO
Basal level correct (perfect scores on 2 consecutive items): YES / NO
Ceiling level correct (4 consecutive scores of 0): YES / NO
Sum of item scores correct: YES / NO
Gave credit for items below basal level: YES / NO
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 item 4 or 5): YES / NO / N.A.

12. SYMBOL SEARCH

1. Completion time recorded: Y / N
2. Number correct recorded: Y / N
3. Number incorrect recorded: Y / N
4. Total Raw Score recorded: Y / N
5. Used black lead pencil: Y / N
6. Practice items completed: Y / N
7. Completion time 120" or less: Y / N
8. Number correct wrong: Y / N
9. Number incorrect wrong: Y / N
10. Total Raw Score wrong: Y / N
11. Sample items completed: Y / N

13. LETTER-NUMBER SEQUENCING

		No. Resp. Recorded	Trial score not recorded	Trial score incorrect	Item score not recorded	Item score incorrect
1.	L-2 (2-L)					
	6-P (6-P)					
	B-5 (5-B					
2.	F-7-L (7-F-L)					
	R-4-D (4-D-R)					
	H-1-8 (1-8-H)					
3.	T-9-A-3 (3-9-A-T)					
	V-1-J-5 (1-5-J-V)					
	7-N-4-L (4-7-L-N)					
4.	8-D-6-G-1 (1-8-6-D-G)					
	K-2-C-7-S (2-7-C-K-S)					
	5-P-3-Y-9 (3-5-9-P-Y)					
5.	M-4-E-7-Q-2 (2-4-7-E-M-Q)					
	W-8-H-5-F-3 (3-5-8-F-H-W)					
	6-G-9-A-2-S (2-6-9-A-G-S)					
6.	R-3-B-4-Z-1-C (1-3-4-B-C-R-Z)					
	5-T-9-J-2-X-7 (2-5-7-9-J-T-X)					
	E-1-H-8-R-4-D (1-4-8-D-E-H-R)					
7.	5-H-9-S-2-N-6-A (2-5-6-9-A-H-N-S)					
	D-1-R-9-B-4-K-3 (1-3-4-9-B-D-K-R)					
	7-M-2-T-6-F-1-Z (1-2-6-7-F-M-T-Z)					

Starting point correct (item 1): YES / NO
Ceiling level correct (failure of all 3 trials of an item): YES / NO / N/A
Sum of item scores correct: YES / NO

WAIS-III ERROR TABLE

	Total Number of Errors	Errors Not Including Failure to Record
Picture Completion		
Vocabulary		
Digit Symbol-Coding		
Digit Symbol- In Learn		
Digit Symbol- Copy		
Similarities		
Block Design		
Arithmetic		
Matrix Reasoning		
Digit Span		
Information		
Picture Arrangement		
Comprehension		
Symbol Search		
Letter-Number Seq		

Totals: _____ _____