An Evaluation of the Efficacy of a Compu-Check-Form for Reducing Computational and Clerical Errors on WAIS-R Protocols

Christine Hodgins ©
Lakehead University

Thesis Supervisor: Anthony P. Thompson, Ph.D. Ç. Psych.

Thesis submitted as partial fulfillment of the requirements for the Master of Arts Degree in Clinical Psychology at Lakehead University, Thunder Bay, Ont.

National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-315-86179-7

Canada

Acknowledgements

I gratefully acknowledge Dr. Tony Thompson for the guidance and encouragement he provided me throughout this project. His socratic style of supervision allowed me to truly learn from my experience. I also thank Dr. E. Rawana for being the second reader for this thesis. In addition, the students and practitioners who volunteered to participate in this study were very much appreciated. Finally, a special acknowledgement goes to The Psychological Corporation for granting permission for Wechsler Adult Intelligence Test - Revised protocols to be duplicated for the purposes of this research.

Table of Contents

List of Tables

List of Appendices

Abstract

A systematic procedure for checking computational and clerical components of scoring (Compu-Check-Form; CCF) was developed and evaluated for the Wechsler Adult Intelligence Scale - Revised (WAIS-R; Wechsler, 1981). Senior undergraduates and M.A. graduate students were trained to score fictitious WAIS-R protocols. Sixty percent of the students made errors and approximately 30% of the protocols contained errors. These errors frequently resulted in IQ discrepancies. Most IQ inaccuracies were small, although 10% of the summary IQs on protocols with errors deviated between 4 and 12 points. A subsample of the students were also trained to use the CCF. Changes in error rates and corrections to summary IQs also supported the utility of the CCF. In a field trial, 6 of 7 practitioners who used the CCF detected errors on 15 of 47 WAIS-R protocols selected from their clinical files. Methodological issues and implications of the results are discussed.

An Evaluation of the Efficacy of a Compu-Check-Form for Reducing Computational and Clerical Errors on WAIS-R Protocols

The Wechsler Intelligence Scales are used extensively with adults and children (Kaplan & Saccuzzo, 1989; Piotrowski & Keller, 1989) and have recently been reported the most frequently and widely used tests with adolescents (Archer, Maruish, Imhof, & Piotrowski, 1991). The scales are considered to be well constructed, highly reliable, valid measures of intelligence (Kaplan & Saccuzzo, 1989; Sattler, 1990) and are viewed by some as the "standard for assessment of intelligence" (Archer et al., 1991, p. 250). Slate and Jones (1990a) noted that the very respectable psychometric properties of the Wechsler scales are based upon the assumption of administration and scoring accuracy. However, research with various Wechsler Intelligence Scales (Wechsler, 1949, 1955, 1974, 1981) has demonstrated that examiners commit numerous errors that detract from the reliability and validity of test results. Similar kinds of examiner errors have been identified for each of the Wechsler tests. These errors can be categorized into three general types: scoring errors, administration errors, and computational and clerical errors.

Scoring Errors

The assignment of incorrect point values to verbal responses on the Comprehension, Similarities, and Vocabulary subtests of the Wechsler scales has long been recognized as a major source of examiner error (Brannigan, 1975; Bradley, Hanna, & Lucas, 1980; Miller & Chansky, 1972; Miller, Chansky, & Gredler, 1970; Slate & Jones, 1989; Slate, Jones, Coulter, & Covert, 1992). It is thought that items on these subtests are susceptible to error because of the semi-subjective nature of response evaluation (Slate & Hunnicutt, 1988; Slate & Jones, 1990b, 1990c). Franklin, Stillman, Burpeau, and Sabers (1982) noted that examiners deficient in the judgement skills required for accurate scoring introduce error to each test administration. Some contend that response scoring guidelines are inadequate and have advised that the scoring criteria in the Wechsler manuals need to be more clearly defined (Brannigan, 1975; Miller & Chansky, 1972; Slate & Chick, 1989; Slate & Hunnicutt, 1988; Slate & Jones, 1990c; Wagoner & Slate, 1988).

Some scoring errors, such as crediting items after the cutoff criterion has been reached, are clear violations of the standard scoring procedures described explicitly in the test manuals. Miller and Chansky (1972) and Miller et al. (1970) noted that scoring errors on the Coding subtest of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949) occurred when examiners neglected to scrutinize subjects' responses, and simply assigned a point to each response attempt. Hajzler (1987) reported that Digit Span was substituted for the Comprehension subtest to derive Verbal IQ scores for Wechsler Intelligence Scale for Children - Revised (WISC-R; Wechsler, 1974) protocols. The only permissible substitution for WISC-R Performance IQ is Mazes for Coding. Other procedural violations found to result in inaccurate scores include use of incorrect formulas to prorate scaled scores (i.e. multiplying by fourth-fifths rather than by five-fourths) and prorating scaled scores subsequent to the arbitrary exclusion of valid subtests (Miller & Chansky, 1972; Miller et al., 1970).

Administration Errors

Administration errors occur when examiners fail to adhere to the standardized procedures for test administration which are made explicit in the test manuals (Moon, Blakey, Gorsuch, & Fantuzzo, 1991). Failure to record examinees' responses, questioning examinees where prohibited, and failure to question when required occur frequently (Conner & Woodall, 1983; Cummings & Moscato, 1982; Moon, Fantuzzo, & Gorsuch, 1986; Slate & Jones, 1990b; Slate & Hunnicutt, 1988; Wagoner & Slate, 1988; Warren & Brown, 1973). Slate et al. (1992) emphasized that recording responses accurately is necessary for verification of scoring accuracy and that a written account of responses is important for subsequent clinical interpretation. Other administration errors that have been reported are improper termination (discontinuance) of subtests, failure to record response times on timed subtests, and reciting digits at the wrong rate for the Digit Span subtest (Franklin et al., 1982; Moon et al., 1991; Slate & Jones, 1990a, 1990b, 1990c; Wagoner & Slate, 1988; Warren & Brown, 1973).

Computational and Clerical Errors

Computational and clerical errors are thought to be due to carelessness (Levenson, Golden-Scaduto, Aiosa-Karpas, & Ward, 1988; Slate & Chick, 1989; Slate & Jones, 1990b, 1990c; Slate & Hunnicutt, 1988). Cummings and Moscato (1982) contend that these errors are especially intolerable because they can easily be detected by rechecking protocols. Simple mistakes in addition of raw and scaled scores have been found to occur most frequently (Cummings & Moscato, 1982; Miller & Chansky, 1972; Sherrets, Gard, & Langner, 1979; Slate & Chick, 1989; Wagoner & Slate, 1988). Mathematical errors resulting in inaccuracies for chronological age have also been reported (Beasley, Lobasher, Henley, & Smith, 1988; Hajzler, 1987; Sherrets et al., 1979; Slate & Jones, 1990b; Slate, Jones, Murray, & Coulter, 1993; Warren & Brown, 1973). The results of several studies indicated that examiners made clerical errors transferring subtest scores to the summary section of the record form, converting raw scores to scaled scores, and deriving IQ scores from the tables in the manual (Beasley et al., 1988; Miller & Chansky, 1972; Miller et al., 1970; Sherrets et al., 1979). Two studies reported blatant clerical oversight in which the scoring of some subtests was omitted (Miller & Chansky, 1972; Miller et al., 1970). Failing to credit nonadministered 'passed' items, circling the wrong point value on Performance subtest items, assigning incorrect point values to objectively scored items (e.g., Arithmetic subtest items), and failing to prorate scores when required have also been reported (Hajzler, 1987; Miller & Chansky, 1972; Miller et al., 1970; Sherrets et al., 1979; Wagoner & Slate, 1988).

Prevalence of Examiner Errors

Errors made by examiners appear to occur at a high rate. Slate and Jones investigated errors on WISC-R and Wechsler Adult Intelligence Scale - Revised (WAIS-R; Wechsler, 1981) protocols in a series of studies with graduate students and practitioners. Their findings, summarized in Table 1, show that students and practitioners made errors on almost all protocols but, on average, practitioners made more errors per protocol than students. Slate and his colleagues maintained that the high error rates among practitioners resulted from examiners who tended not to record responses to test items. When failures to record item responses were not

Table 1

Error Rates on WISC-R and WAIS-R Protocols Across Various Studies

Study	Sample	Protocols		
		n	% with errors	M (SD)[1]
		WISC-R		
Slate & Jones (1990c)	Stud.	217	100	11.3 (15.1)
Slate et al. (1992)	Prac.	56	100	38.4 (29.3)
		WAIS-R		
Slate & Jones (1990a)	Stud.	149	97	8.0 (6.9)
Slate & Jones (1990b)	Stud.	180	98	8.8 (5.6)
Slate et al. (1993)	Prac.	50	100	36.9 (22.6)

Note. Stud. · Students, Prac. · Practitioners.

1Mean (Standard Deviation) errors per protocol.

included as errors on protocols, mean errors per WISC-R protocol became similar for practitioners and graduate students alike (e.g., 8.7 for practitioners for Slate et al., 1992). However, for the same analysis with the WAIS-R, practitioners still made almost two times as many errors as graduate students (e.g., 15.4 for practitioners in Slate et al, 1993). Slate's research also revealed that practitioners made two types of errors that students did not. These were neglecting to record raw subtest scores in the designated place and omitting days in the test age calculation (Slate et al., 1992; Slate et al., 1993).

Several researchers have focused upon computational and clerical errors. Hajzler (1987) found errors on 64% of WISC-R protocols obtained from the files of a psychoeducational service organization. Similarly, Sherrets et al. (1979) reported 46.5% of 200 WISC and WISC-R protocols randomly selected from psychiatric and school records contained computational and clerical errors. Over 89% (N = 39) of the examiners who scored these protocols made at least one error. As part of a longitudinal study of children who received early treatment of phenylketonuria, Beasley et al.

(1988) checked 457 (184 WISC, 273 WISC-R) protocols scored by clinical and educational psychologists from over 100 agencies in the United Kingdom. Twenty-four percent of the protocols, which were checked by a custom written computer program, contained errors.

Levenson et al. (1988) found that computational and clerical errors on WISC-R protocols occurred significantly more for PhD. school psychologists (74%) compared with Master's level school psychometrists (45%), and graduate school psychology interns (50%). Similarly, Ryan, Prifitera, and Powers (1983) reported doctoral level psychologists were more prone to make clerical errors on WAIS-R protocols than psychology interns. Some have speculated that employment related stressors such as time pressure and large caseloads might account for greater error rates found among professional psychologists (Levenson et al., 1988; Slate & Hunnicutt, 1988).

Impact of Examiner Errors

It has been suggested that variability in scoring can be used as an estimate of examiner error with the Wechsler scales. For example, examiners have scored identical protocols that were fabricated by researchers

(Bradley et al., 1980; Miller & Chansky, 1972; Miller et al., 1970) or were reproductions of actual protocols (Oakland, Lee, & Axelrad, 1975; Ryan et al., 1983). To evaluate the impact of examiner error, the standard deviation (SD) of IQ scores was compared to the standard error of measurement (SEM). Most studies found that the SD of Full Scale IQ scores was smaller than the SEM (Miller & Chansky, 1972; Miller et al., 1970; Oakland et al., 1975; Ryan et al., 1983). Similar results were reported by Warren and Brown (1973) for WISC protocols administered to actual clients, and by Franklin et al. (1982) for Wechsler Adult Intelligence Scale (WAIS) protocols administered to confederates trained to give standard responses to test items.

Oakland et al. (1975) correctly noted that SDs, as a measure of error, reflect different sources of error than SEMs in the Wechsler manuals. Standard Error of Measurements are based on internal consistency and stability coefficients that reflect error variance due to content and time sampling whereas SD reflects imperfect scoring reliability (Hanna, Bradley, & Holen, 1981; Ryan et al., 1983; Slate & Chick, 1989; Slate & Hunnicutt, 1988). However, Oakland and his colleagues concluded that examiner error had an insignificant impact on IQ. It seems more appropriate to emphasize that examiner error is an additional source of error. Therefore, SEMs underestimate the range of error (Hanna et al., 1981; Slate & Hunnicutt, 1988).

Hanna et al. (1981) have estimated an error index that considers examiner and internal error. They estimated a composite SEM of 6.60 for the WISC-R Full Scale IQ. Slate and Hunnicutt (1988) pointed out this estimate is twice as great as the average SEM of 3.19 reported in the WISC-R test manual. As all of the Wechsler subtests are susceptible to examiner error, similar logic applies to the measurement error for subtest scaled scores (Franklin et al., 1982; Miller and Chansky, 1972; Miller et al., 1970; Slate & Chick, 1989; Wagoner & Slate, 1988). Hence, it has been suggested that examiners exercise extreme caution when making clinical interpretations based on the scatter of individual subtest scaled scores (Bradley et al., 1980; Cummings & Moscato, 1982; Franklin et al., 1982).

The impact of examiner error has also been examined by discrepancy scores (e.g., differences between IQ scores on protocols with errors and corrected protocols). Beasley et al. (1988) found discrepancies on 28% of 457 WISC and WISC-R protocols. Discrepancies between computed and 'true' scores on WISC-R protocols ranged from -19 to +15 for Verbal IQ, -6 to +22 for Performance IQ, and -11 to +14 for the Full Scale IQ scores. Comparable ranges were reported for the WISC. Five percent of all protocols had IQ errors that exceeded 5 points. Similarly, Cummings and Moscato (1982) found discrepancies on WISC-R protocols that ranged from 1 to 14 points for Verbal IQ, 1 to 13 points for Performance IQ, and 1 to 15 points for Full Scale IQ scores. The ranges of IQ errors reported by Beasley et al. (1988) and Cummings and Moscato (1982) are greater than those reported by Slate et al. (1992). Slate and his colleagues found that 81.5% of WISC-R protocols had IQ score errors, but no error exceeded 4 IQ points. Research with the WAIS-R (Slate and Jones, 1990a, 1990b; Slate et al., 1993) revealed Full Scale IQ score discrepancies that ranged from 1 to 10 points. IQ score errors were found on 54% (Slate et al., 1993), 72% (Slate & Jones, 1990a), and 81% (Slate & Jones, 1990b) of protocols.

IQ score inaccuracies of the magnitude found in many studies could result in individuals being misclassified (i.e., intellectually deficient) and misplaced in special programs (Cummings & Moscato, 1982; Franklin et al., 1982; Miller et al., 1970; Slate & Hunnicutt, 1988). In other cases, errors might result in the exclusion of individuals from special classes, along with expectations for their success in the regular stream (Beasley et al., 1988; Hajzler, 1987; Levenson et al., 1988; Miller & Chansky, 1972; Slate & Chick, 1989; Warren & Brown, 1973).

Reducing Examiner Error

Slate and Hunnicutt (1988) suggested inadequate training may be the root cause of examiner error. Training courses for assessment are typically comprised of demonstrations of test administration, discussions of administration and scoring procedures, and several practice administrations (Slate & Jones, 1990c). A national survey of course instructors found practice administration to be a principal component of most courses (Oakland & Zimmerman, 1986). The survey indicated that, on average, 6.7 practice administrations of the WISC-R are required, followed by 3.9 practice administrations of the WAIS-R. Slate, Jones, and Murray (1991) noted that this approach is based on two assumptions: 1) students will acquire competency in testing through practice, and 2) students require fewer WAIS-R administrations because skills acquired through practice with the WISC-R will transfer to the WAIS-R.

Slate et al. (1991) tested both of these assumptions. Students' accuracy for initial administration of the WAIS-R was examined to determine the effects of five and ten practice administrations with the WISC-R. The results indicated that increasing the number of practice administrations with the WISC-R did not significantly improve students' accuracy of initial administrations of the WAIS-R. Further, it was suggested that some skills acquired with the WISC-R contributed to errors on the WAIS-R. Specifically, the process for converting raw scores to scaled scores on the WISC-R was inappropriately used on the WAIS-R. Analyses of the effects of practice administrations with the WAIS-R showed no significant reduction in errors across five administrations, but a small statistically significant decrease in errors was found for ten administrations. Slate et al. (1991) reported that the decrease was due to the reduction of failure to record errors. When these errors were excluded from their analyses they found no effect across ten administrations and a significant increase in errors across five administrations. Several other studies of the Wechsler scales have reported similar results (Patterson, Slate, Jones, & Steger, 1991; Slate and Jones, 1990c; Slate, Jones, & Covert, 1992; Warren and Brown, 1973).

Conner and Woodall (1983) developed a WISC-R error checklist to provide structured feedback to students after each of fifteen test administrations. The checklist included selected administration errors (i.e., failure to record verbal responses, failure to question when required) and computational and clerical errors. Subtests with response scoring errors were also identified. The results indicated that the checklist reduced the frequency of administration errors but the frequencies of response scoring and computational and clerical errors were unaffected. A version of Conner and Woodall's (1983) checklist was also used by Slate and his colleagues for their research. They advised giving students immediate feedback during practice administrations to reduce the likelihood of repeating errors. The disadvantage of this procedure is that it interrupts testing (Slate et al., 1992).

Slate and Jones (1989) argued that more emphasis should be placed on instruction to improve students' assessment skills. They found that students who were informed about common errors and given explicit rules to avoid them, prior to administering the WISC-R, made significantly fewer errors across seven administrations than students who only received feedback. A competency-based training model (MASTERY model), originally proposed by Fantuzzo (1984) also focuses on instruction. The model provides a method of systematically evaluating, to a 90% accuracy criterion, students' knowledge and performance of standard administration procedures for the WISC-R (Fantuzzo, Sisemore, & Spradlin, 1983) and WAIS-R (Blakey, Fantuzzo, & Moon, 1985; Fantuzzo & Moon, 1984; Moon et al., 1986). More recently the MASTERY model has been extended to include response scoring for the WAIS-R (Blakey, Fantuzzo, Gorsuch, & Moon, 1987). Students trained with this model have reached the criterion level after ten to fifteen hours of training and only two to three practice administrations.

Most error remediation research has focused on response scoring and administration errors. Conner and Woodall (1983) suggested response scoring errors should be of the most concern because they occur so frequently. However, Franklin et al. (1982) found no direct relationship between the frequency of response scoring errors and IQ score inaccuracies. Also, error frequency can be misleading because of the different opportunities for various errors to occur. The possibility of a response scoring error exists for every test item administered but a computational error, such as the miscalculation of age, can occur only one time per protocol. Cummings and Moscato (1982) found that administration and response scoring errors accounted for 85% of the total errors on WISC-R protocols. However, less frequent computational and clerical errors, particularly mistakes in simple addition, had the greatest impact on the accuracy of IQ scores.

The purpose of the present research was to develop and evaluate a procedure for checking computational and clerical components of scoring the WAIS-R protocol.

Some of the explicit rules for avoiding computational and clerical errors, as suggested by Slate and Jones (in press) were incorporated. The following research questions were addressed: 1) What is the nature and frequency of computational and clerical errors? 2) What is the impact of errors on accuracy of IQ scores? 3) Does use of systematic checking procedures significantly reduce errors?

Study 1

Method

Compu-Check-Form. A Compu-Check-Form (CCF) for the WAIS-R was developed as an aid for computing summary scores from raw data. The CCF (see Appendix C) is comprised of simple arithmetic algorithms and clerical checking procedures presented in a clear and well organized format. Some procedures (e.g., Age Calculation Check) were based on strategies described by Slate and Jones (in press) and others were developed by the researchers. The checking procedures and layout are as follows.

Age Calculation Check. A reminder of year-month and month-day equivalents used for computing test age appears over a computation box. The computation box is designed to check that test age was calculated correctly. As on the WAIS-R record form, the test age is calculated by subtracting the subject's birth date from the test date. On the CCF, the birth date is then added to the calculated test age to compute a test date. If the test age is calculated correctly, the computed test date should correspond with the actual test date. As a final step, test age on the record form is required to match test age on the CCF.

Raw Score Checks. A computation box for calculating raw score totals is provided for each subtest. To calculate a subtest total, the examiner is required to count the number of items scored 0 (items not administered above the ceiling are scored 0). The number of 0 point responses is then entered in the appropriate square in the computation box. This operation is repeated for items scored 1 point and where applicable for items scored 2 points, 3 points etc. Frequency counts (i.e., numbers in the squares) are summed and should equal the total number of items for that subtest (also provided on the CCF). Next, the frequency counts are multiplied by the item score values and the product is recorded in a circle on the right side of the computation box. The products in the circles are added to compute the raw score subtest total. Subtest totals obtained using the CCF are compared with corresponding subtest totals recorded on the summary sheet of the WAIS-R Record Form.

Scaled Score Checks. Tables for converting raw scores to scaled scores appear to the right of each computation box. A separate table is provided for each subtest to avoid errors associated with reading from the wrong column of the Table of Scaled Score Equivalents on the WAIS-R record form. After each subtest total is computed, the raw score is converted to a scaled score using the adjacent table. The obtained scaled score is then entered in an oval to the right of the conversion table and compared with the scaled score for the same subtest recorded on the summary sheet of the WAIS-R Record Form. This procedure is repeated for each subtest. Colour codes are used to make Verbal and Performance subtest scaled scores easily distinguishable when they are subsequently transferred to calculate the Verbal and Performance scaled score totals.

Verbal Score Calculation Check. The scaled scores for Verbal subtests recorded in yellow ovals are transferred to corresponding yellow ovals in the Verbal Score computation box. The scores in the ovals are then summed to obtain the Verbal scaled score total. The Verbal score is recorded in a yellow hexagon and compared with the Verbal Score recorded on the summary sheet of the WAIS-R Record Form.

Performance Score Calculation Check. The procedure is the same as that described for checking the Verbal Score except that the individual Performance subtest scaled scores are transferred to corresponding blue ovals in a computation box and then the total score is highlighted in a blue hexagon.

Subjects The sample consisted of forty university psychology students (33 female, 7 male). Thirteen subjects were graduate students enrolled in the first year of a two year Master's program in Clinical

Psychology. The remaining 27 subjects were enrolled in 4th year Honour's psychology courses. Prior to participating in this study, the students had not administered or scored an individually administered intelligence test, nor had they received classroom instruction regarding these procedures.

Materials

WAIS-R Protocols. Permission was obtained from The Psychological Corporation to duplicate protocols for research purposes (Appendix B). Five WAIS-R protocols were fabricated. For each protocol, fictitious item responses for the eleven subtests were recorded on expanded WAIS-R record forms. Appropriate scores were assigned to each item response and the point values were recorded in the spaces allotted. The protocols were constructed to sample a range of Full Scale IQ classifications. For each of the 5 protocols, a fictitious subject was identified by letter only (i.e., A, B etc.). This information, as well each fictitious subject's birth date and the date tested were recorded on the summary page of the record form. The original fabricated protocols were duplicated to ensure subjects received identical record forms.

<u>Procedure</u> Ethical approval for this research was granted by Lakehead University's Ethics Advisory Committee to the Senate Research Committee (Appendix A). Subjects were apprised of the purpose of the study and volunteered to participate (Appendix C). The study consisted of a training phase and a testing phase. The first 30 subjects were randomly assigned to one of two training conditions. Fifteen subjects were trained in standard scoring procedures only, and 15 subjects were trained in standard scoring procedures <u>and</u> use of the CCF. An additional 10 subjects were subsequently trained in standard scoring procedures and use of the CCF for a total of 25 subjects in this condition. Subject categorization as graduate versus undergraduate was not significantly associated with training group assignment. In both training conditions the subjects were instructed individually or in small groups consisting of no more than 3 subjects. The total training time for standard scoring procedures was approximately 30 minutes for both groups. Subjects trained to use the CCF received an additional 20

minutes of training time that was devoted to CCF procedures only.

<u>Standard Scoring Procedures Training.</u> Subjects were provided with a WAIS-R test manual and fictitious WAIS-R protocol (Practice A). Instructional aids used by the researcher included a test protocol printed on transparencies that was identical to the protocol given the subjects, and an overhead projector. To begin training, a brief introduction was given to familiarize the subjects with the record form. Following this, through verbal instruction and visual demonstration, subjects were trained to: 1) calculate the fictitious subject's test age, 2) add item scores to obtain a total raw score for each subtest, 3) transfer each subtest raw score total to the summary page of the record form, 4) use the tables on the record form to convert raw subtest scores to scaled subtest scores, 5) sum the scaled scores of the Verbal subtests to obtain a Verbal score, 6) sum the scaled scores of the Performance subtests to obtain a Performance score, 7) sum the Verbal and Performance scores to obtain the Full Scale score, 8) use the tables in the manual to

convert the Verbal score to a Verbal IQ score, 9) use the tables in the manual to convert the Performance score to a Performance IQ score, and 10) use the tables in the manual to convert the Full Scale score to a Full Scale IQ score.

A practice session immediately followed the instructional stage of training. Subjects were given a different fictitious WAIS-R protocol (Practice B) to complete. They were monitored to determine if they could complete the protocol using the standard scoring procedures. While monitoring subjects, procedural questions were answered but no other help was given. Once the protocols were completed, the subjects were provided feedback which consisted of a summary page from the record form containing the correct calculated test age and summary scores (i.e., raw scores, scaled scores, IQ scores).

Compu-Check-Form Training. Subjects in the CCF group received scoring procedures training that was the same as that given subjects in the standard training group. However, after completing practice protocol B, they were given a CCF to check their scoring. Verbal instructions supplemented by visual demonstration on an overhead projector (i.e., transparencies of the CCF), were used to explain the use of the CCF. Instructions proceeded in a piecemeal fashion with subjects completing each section of the CCF before proceeding to instructions for the subsequent section. If the CCF procedure revealed a scoring discrepancy, subjects were instructed to circle the discrepancy on the record form and record the change next to it. Also, subjects were instructed to follow through with changes, where required, to Full Scale and IQ scores which were not directly checked with the CCF. Subjects were monitored and procedural questions were answered. Following completion of the checking procedure, feedback was given to subjects via a transparency of the CCF with correct scoring and oral feedback for the summary IQ scores.

Testing Phase. Testing took place immediately following training for all subjects. A separate room was provided for each subject tested. Each subject was given 3 fictitious WAIS-R protocols (protocols A, B, and C). In addition, each subject in the CCF group received three CCFs. All subjects received the same standard instructions to score the protocols as they had been trained and to score the protocols carefully. The protocols were completed in a predetermined order that was the same for all subjects. A time limit was also imposed and subjects were instructed that they had "up to 55 minutes to complete the scoring".

Treatment of Data The protocols completed during the testing phase were examined to determine the frequency and type of errors. For the purpose of analysis, 10 types of errors were defined. Definitions were similar to those applied in previous studies (e.g., see Hajzler 1987; Moscato & Cummings, 1982). The error types and definitions were as follows.

1. Age Error: Subtraction error calculating the test age.
2. Raw Score Error: Error summing the raw score of a verbal or performance subtest.
3. Transferring Error: Clerical error transferring the raw score total of a verbal or performance subtest to the summary table of the record form.
4. Subtest Scaled Score Error: Obtaining an incorrect

scaled score for a Verbal or Performance subtest. An incorrect scaled score resulted during the conversion process (i.e., misread from the Table of Scaled Score Equivalents) or from a previous error (i.e., incorrect raw score total).

5. Verbal Score Error: Obtaining an incorrect Verbal Score. This error resulted from incorrectly summing the scaled scores of the verbal subtests or from a previous error.

6. Performance Score Error: Obtaining an incorrect Performance Score. This error resulted from incorrectly summing the scaled scores of the performance subtests or from a previous error.

7. Full Scale Score Error: Obtaining an incorrect Full Scale Score. This error resulted from incorrectly summing the Verbal and Performance Scores or from a previous error.

8. Verbal IQ Error: Obtaining an incorrect VIQ. This error resulted from incorrectly converting the Verbal Score to an IQ score or from a previous error.

9. Performance IQ Error: Obtaining an incorrect PIQ. This error resulted from incorrectly converting the Performance Score to an IQ score or from a previous

error.

10. Full Scale IQ Error: Obtaining an incorrect FSIQ. This error resulted from incorrectly converting the Full Scale Score to an IQ score or from a previous error.

Errors were also categorized as initial or consequent errors. An initial error resulted directly from an error made by the subject who scored the protocol. A consequent error was a subsequent error that was generated by an initial error(s). Total errors equalled the sum of initial and consequent errors.

Results All 15 subjects who received training in the standard scoring procedures completed 3 protocols in the testing phase for a total of 45 protocols. One subject required the full 55 minutes to score 3 protocols. All other subjects in the group scored the protocols within 35 to 45 minutes. Sixty-eight protocols (record form and CCF) were completed by subjects in the CCF group. Seven subjects only partially completed the third protocol within the allotted time and these incomplete data were not included in the analyses.

The absolute frequency and relative frequency of errors made on the WAIS-R protocols by training condition are presented in Table 2. Table 2 shows that subjects in the standard training condition and in the CCF condition prior to applying the checking procedure (pre-CCF), made errors in all but one of the error categories. The number of errors per protocol ranged from 0 to 7 for the standard training group (M = .93, SD = 1.97) and from 0 to 10 for the pre-CCF subjects (M = 1.44, SD = 2.63). Individual subject means ranged from 0 to 4.3 errors per protocol for the standard training condition, and from 0 to 8 for the pre-CCF condition. The error rates for both groups were similar in most instances, although the pre-CCF group appeared to have relatively more age, scaled score errors, Performance total errors, and higher overall error rates. Twelve of 42 (28.6%) errors made by the standard scoring group were initial errors compared with 36 of 100 (36.0%) made by the pre-CCF group. The most frequent initial errors for both groups were for calculating Age, Raw Score totals, and deriving Subtest Scaled Scores. Initial errors generated an average of 2.5 consequent errors per protocol for the standard

Table 2

Absolute and Relative Frequency of Errors on WAIS-R Protocols by Training Condition

	Training Group[1]					
	Standard		Compu-Check-Form			
			Pre		Post	
Error Type	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.
1. Age	4	.09	12	.18	6	.09
2. Raw Score	3	.06	7	.10	0	.00
3. Transferring	0	.00	0	.00	0	.00
4. Scaled Score	5	.11	17	.25	1	.01
5. Verbal Score	5	.11	9	.13	1	.01
6. Perf. Score	1	.02	8	.12	0	.00
CCF Subtotal	18	.40	53	.78	8	.12
7. FS Score	7	.15	15	.22		
8. VIQ	[illegible]	.13	12	.18		

Table 2 Continued

	Training Group[1]					
	Standard		Compu-Check-Form			
			Pre		Post	
Error Type	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.
9. PIQ	2	.04	8	.12		
10. FSIQ	9	.20	12	.18		
Total	42	.93	100	1.47		

<u>Note.</u> Abs. · absolute frequency; Rel. - absolute frequency divided by number of protocols.

[1]Fifteen subjects in the Standard training condition, 25 subjects in the Compu-Check-Form condition.

scoring procedures group and 1.9 consequent errors per protocol for the pre-CCF group.

Evaluation of the CCF focused primarily on those error categories that were directly checked with the CCF. This was necessary as most subjects who used the CCF seemed to forget the additional instruction to follow through with changes on protocols (i.e., recalculation of Full Scale Score and IQ scores). The data in Table 2 support the contention that the CCF reduces computational and clerical errors. Specifically, subjects in the CCF group seem to reduce their total errors after using the checking procedure. Also the rate of total errors post-CCF seems smaller than the error rate for subjects who did not have the benefit of CCF training. The pattern of initial errors, not shown in Table 2, also supported the effectiveness of the CCF.

To statistically assess the impact of the CCF, the number of subjects making errors was determined for all conditions. Eight (53%) subjects in the standard training group and 16 (64%) subjects in the CCF group made errors prior to applying the CCF. A Pearson Chi-square test of association between training group (standard vs. pre-CCF) and subject categorization (errors vs. no errors) was not significant; $X2$ $(1, N = 40) = .44$, $p < .50$. Subsequent to checking, 6 (24%) subjects in the CCF group made errors. For this data the test of association between training group (standard vs. post-CCF) and subject categorization (error vs. no error) was in favour of CCF effectiveness; $X2$ $(1, N = 40) = 3.55$, $p < .06$. A test for the significance of a difference between two correlated proportions (Ferguson, 1971) was used to statistically assess the pre- and post-CCF error rate. The result was significant in favour of a reduction in proportion of subjects making errors subsequent to use of the CCF ($z = -3.16$, $p < .01$). It should be noted that because of the small sample size, this test statistic may not be a good approximation of the normal curve.

The number of protocols containing errors checked by the CCF also supported its effectiveness. In the standard training condition, 20% of the protocols contained errors. In the CCF condition, 34% of the protocols contained errors prior to checking, compared with only 10% of the protocols after applying the CCF.

The data could not be analyzed statistically because more than one observation (protocol) was produced by each subject.

Protocols with errors were examined to determine the impact on IQ scores. Discrepancies between calculated and 'true' IQ scores for the standard training group and pre-CCF group combined are presented in Table 3. Eighteen (50%) of the protocols had Verbal IQ (VIQ) errors, 10 (28%) had Performance IQ (PIQ) errors, and 21 (58%) Full Scale IQ (FSIQ) errors. Eighty-nine percent of VIQ and PIQ, and 86% of FSIQ scores were within 3 points of the actual scores. Mean discrepancies were 1.78 (<u>SD</u> = 2.80) for VIQ, 1.03 (<u>SD</u> = 2.41) for PIQ, and 1.33 (<u>SD</u> = 1.59) for FSIQ.

Six protocols scored and checked by 5 subjects who followed through with changes to summary scores not included on the CCF were examined for IQ errors. All protocols contained at least 1 IQ error before the CCF was used. Four protocols had VIQ errors, 3 had PIQ errors, and 5 protocols had FSIQ errors. Discrepancies ranged from 1 to 3 points for VIQ, 1 to 4 points for PIQ, and 1 to 3 points for FSIQ scores. After the CCF was used there were no IQ errors.

Table 3

Discrepancies of IQ Scores from the 'True' Scores for the Standard Training Group and Pre-CCF Group, Combined

IQ Points	VIQ	PIQ	FSIQ
0	18	26	15
1	4	2	9
2	5	3	4
3	5	2	4
4			1
5			
6			
7			
9			
11			
12			
Total	36	36	36

Study 2

Method

Subjects The sample consisted of 10 practitioners (6 Ph.D., 3 M.A., 1 advanced level M.A. student) working in clinical settings. Eight subjects were male and 2 were female. All worked in hospital settings and had a minimum of 1 year of experience administering the WAIS-R to clients. Subjects were apprised of the purpose of the study and volunteered to participate (Appendix C).

Procedure Subjects were trained to use the CCF and subsequently used CCFs to check WAIS-R protocols that had previously been administered in their clinical practice. They were also asked to give their impressions of the CCF on a brief questionnaire (see Appendix B).

Subjects' work schedules necessitated flexible procedures for CCF training. Seven subjects received training in one, twenty minute group session. During group training each subject was given a CCF and was verbally instructed in its use. Three subjects were instructed individually. Training time was approximately thirty minutes. The additional time allowed for practice in use of the CCF. All subjects were advised to contact the researcher if they experienced difficulties with subsequent use of the CCF.

Subjects were required to select 10 WAIS-R protocols from archival client files. To ensure confidentiality, the identity of clients whose protocols were selected was known only to the subjects. Protocols were checked at the subjects' convenience but were completed within three weeks from training. Subjects were informed that they should follow through with checks not included on the CCF. If errors were detected, they were to be noted on the CCF. After checking protocols, subjects answered the opinion questionnaire. To ensure subject confidentiality, no identifying marks were placed on the CCFs or questionnaires. To distinguish among CCFs used by individual subjects, the forms were bundled separately by the clinicians. The individual bundles of CCFs and questionnaires were returned as a group to the researcher for analysis.

Results

Seven subjects returned 47 completed CCFs (range 3-10 protocols per subject). For purposes of analysis, the same categories of errors defined in Study 1 were used, except for Transferring and Full Scale Score errors. Transferring errors were not included because the researchers did not have access to the original protocols to determine if these errors occurred. Full Scale Scores were not directly checked by the CCF. All subjects who detected errors with the CCF followed through with checks but only 1 subject identified Full Scale Score errors. Most subjects indicated the magnitude of IQ discrepancies errors only, and Full Scale Score errors could not be determined from the data provided by subjects on the CCFs.

Six (86%) subjects detected errors on 15 (32%) protocols. The frequency of error types are presented in Table 4. Table 4 shows that practitioners made errors in all but one of the categories. Individual means ranged from 0 to 2.6 errors per protocol. Thirty-four (69%) errors were detected by direct checks with the CCF, and 18 of these were initial errors. Fifteen (31%) errors were detected when subjects followed through with summary score checks not included on the CCF. Only 2 (7%) of these summary errors were initial

Table 4

<u>Frequency of Errors Made by Practitioners that were Detected with the Compu-Check-Form</u>

Error Type	Frequency of Total Errors	Frequency of Initial Errors
1. Age	0	0
2. Raw Score	13	13
3. Transferring		
4. Scaled Score	10	
5. Verbal Score	4	0
6. Performance Score	6	[illegible]
CCF Subtotal	34	18
7. Full Scale Score		
8. Verbal IQ		0
9. Performance IQ	6	2
10. Full Scale IQ	5	0
Total	49	20

errors.

Protocols with errors were examined to determine the impact on IQ scores. As one subject did not follow through with IQ checks on two protocols, the data are based on 13 protocols. Subjects detected IQ errors on 9 protocols. Table 4 shows that most IQ inaccuracies resulted from previous errors. Discrepancies between calculated and corrected IQ scores ranged from 0 to 2 for VIQ, 0 to 11 for PIQ, and 0 to 3 for FSIQ scores. Mean discrepancies were .54 (SD = .88) for VIQ, 1.62 (SD = 3.01) for PIQ, and .62 (SD = .96) for FSIQ scores.

The subjects returned six completed questionnaires. Five subjects indicated that they typically check their scoring. Three reported that they check protocols by repeating their scoring and computations. Two subjects reported that they use a different method for checking their initial computations, but they provided no details about the different methods. Five subjects indicated they would use some CCF techniques, but only two reported that they would personally use the CCF. The CCF techniques subjects preferred were Age and Raw Score checks. Three of four subjects who supervise graduate students recommended use of the CCF for psychology interns. The CCF was also recommended by subjects for file audits and research.

Discussion

The results of this study are consistent with previous findings (Beasley et al., 1988; Cummings and Moscato, 1982; Hajzler, 1987; Levenson et al., 1988; Ryan et al., 1983; Sherrets et al., 1979) that computational and clerical errors occur frequently during scoring of Wechsler intelligence test protocols. Sixty percent of students and most practitioners made errors scoring protocols. Approximately one-quarter to one-third of the protocols contained errors. Students and practitioners made similar types of errors. Most errors resulted from mistakes in simple calculations and clerical procedures and appeared to be due to carelessness. None of the practitioners made age errors, but age errors accounted for approximately 11% of the total errors made by subjects in the standard training and pre-CCF conditions. Most age inaccuracies were for calculations that required the students to regroup year, month, and days. As the students had not scored an intelligence test prior to participating in the present study, they were unaccustomed with the regrouping formula for calculating chronological age. The relatively high rate of age errors among students may have resulted from this lack of experience. Similar to the findings of previous research (Slate et al., 1992, 1993), practitioners made errors that students did not make. One practitioner neglected to calculate an examinee's age on the original protocol and two examiners indicated that they typically do not calculate months and days in the chronological age. These errors were not included in the frequency counts because they were beyond the operational definition of errors used in this study. The errors did not impact on the accuracy of IQ calculations for practitioners in this study, but do violate the standard procedures described in the WAIS-R manual.

Most errors made by students and practitioners were consequent errors. Age errors generated only one consequent error. In that instance, the calculated age was in error by ten years and the wrong tables were used to convert sums of scaled scores to IQ scores.

With the exception of one subtest scaled score error that was 'cancelled out' by a subsequent mistake in addition, all other initial errors affected accuracy of IQ scores. Similarly, all but one initial error detected by the practitioners affected IQ score accuracy.

Mean IQ error from computational and clerical mistakes was small and most of the IQ errors were within 3 points of correct IQ scores. However, 10% of the summary IQs for Study 1 protocols with errors contained substantial IQ errors (4-12 IQ points). This range of IQ discrepancies illustrates the dramatic impact for individual subjects. A similar situation existed for the practitioners in Study 2. Most IQ discrepancies were small, but a discrepancy of 11 IQ points that was reported by one practitioner resulted from a prorating error. Again, this demonstrates the potential impact of a single initial error. Cummings and Moscato (1982) have emphasized the human fallibility of examiners. The results of this study underscore their point and suggest that it would be arrogant for examiners to assume mistakes are not made, or that they are rare and of no clinical significance.

Also, as Slate and his colleagues have emphasized, examiner errors are in addition to internal error. Computational and clerical errors of the type investigated here detract from the reliability and validity of tests scores.

Evaluation of the Compu-Check-Form's utility for reducing computational and clerical errors was promising. In Study 1, fewer subjects made errors after applying the CCF. Changes in error rates were also in favour of the CCF procedure. Reductions in the number of protocols containing computational and clerical error also supported the value of the CCF as an effective checking procedure. In addition, practitioners were able to identify computational and clerical errors in previously scored protocols with the assistance of the CCF.

All subjects were told to score the protocols carefully but, for the standard training group, checking was entirely at each subject's discretion. Anecdotally, subjects in the standard training condition reported checking some components of scoring but they still made errors. Similarly, subjects trained to use the CCF, and therefore alerted to computational and clerical errors, made errors before they applied the CCF. These observations imply that general caution and even knowledge of specific errors are not sufficient to avoid errors. A specific checking procedure seems necessary. This conclusion is supported by previous research (Slate & Jones, 1989) which showed graduate students made fewer errors when they were alerted to errors and given rules to avoid them.

The CCF group initially seemed to make more errors than the standard training group. Two subjects in the CCF condition augmented the total error rates per protocol, although these subjects' data did not entirely account for the pattern. Greater learning demands during training with the CCF may have distracted or fatigued subjects. Also they may have relied more on the checking procedure for accuracy than their initial calculations. The standard training condition and pre-CCF condition were comparable in the proportion of subjects who made errors.

The results of the present research also revealed areas for improvement in the design of the CCF. Six of 8 (75%) of errors in the post-CCF condition were age miscalculations and some students did not complete the age check on the CCF. As previously mentioned, students experienced difficulty with year-month and month-day equivalents. In contrast, experienced practitioners endorsed the age check as a preferred checking procedure. Research with a larger sample of individuals experienced with use of psychological tests is needed for a meaningful appraisal of the age check. A significant problem with the design of the CCF was that it did not provide checks for the Full Scale and IQ scores. Only 5 of 16 students who made errors followed through with changes on protocols. All Full Scale and most IQ score errors resulted from mistakes that were corrected with use of the CCF. It seems highly probable that these errors would have been caught if Full Scale and IQ score checks had been included in the design of the CCF. The absence of IQ errors by the subjects who did follow through supports this contention. However, further validation would be worth while with checks for the Full Scale and IQ scores incorporated into the design of the CCF.

Practitioner feedback about the CCF was encouraging. Five of six practitioners indicated that they typically check their scoring. This suggests that many examiners appreciate that precautions should be taken to avoid errors. While most practitioners indicated they would not personally use the CCF, they were receptive to use of some of the CCF checking techniques. It is likely that practitioners would not routinely use CCFs because of the time involved. However, once they are familiarized with the CCF, practitioners may be able to apply the essential techniques more expediently but equally effectively. Recommended uses for the CCF in clinical settings included supervision of psychology interns and file audits.

The sample of the present study was somewhat unique in that it included undergraduate and graduate students. However, undergraduates were 4th year students and possibly not that different from graduates with whom this type of research is typically conducted. For example, errors similar to those made by graduate students in other studies were observed in Study 1. Nevertheless, additional research with the CCF and other graduate trainees should be pursued. Limitations with the practitioner sample also requires cautious interpretation of the findings. The sample was small and examiners in only two clinical settings were approached to participate. Therefore, the sample may not be representative of examiners working in diverse settings. The sample of practitioners was also heterogeneous in terms of education (i.e., advanced level graduate student to PhD.) and experience (i.e., 1 year to several years). Another potential limitation of the present study was that the protocols that were checked by practitioners were selected by the subjects and errors were self-reported. It is possible that the protocols were not representative and/or that errors were under reported. However, the proportion of protocols with errors made by practitioners in this study was similar to previous research with practitioners (Beasley et al., 1988; Cummings and Moscato, 1982; Hajzler, 1987; Levenson et al., 1988; Sherrets et al., 1979). In spite of the noted limitations, the findings of this research were encouraging and the Compu-Check-Form seems to have passed initial muster. Further development, application, and evaluation of the CCF would be beneficial. Also, many of the CCF WAIS-R checking techniques (e.g., Age, Raw Score checks) could be adapted to check computational and clerical errors on the Wechsler Preschool and Primary Scale of Intelligence - Revised (Wechsler, 1989) or the Wechsler Intelligence Scale for Children - Third Edition (Wechsler, 1991).

References

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, 247-252.

Beasley, M. G., Lobasher, M., Henley, S., & Smith, I. (1988). Errors in Computation of WISC and WISC-R intelligence quotients from raw scores. Journal of Child Psychology and Psychiatry, 29, 101-104.

Blakey, W. A., Fantuzzo, J. W., Gorsuch, R. L., & Moon, G. W. (1987). A peer-mediated competency-based training package for administering and scoring the WAIS-R. Professional Psychology: Research and Practice, 18, 17-20.

Blakey, 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, 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.

Brannigan, G. G., (1975). Scoring difficulties on the

Wechsler intelligence scales. Psychology in the Schools, 12, 313-314.

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, 247-252.

Cummings, J. A., & Moscato, E. M. (1982). WISC-R errors of state certified school psychologists. Unpublished manuscript, Institute for Child Study, Indiana University, Bloomington, IN.

Fantuzzo, J. W. (1984). MASTERY: A competency-based training model for clinical psychologists. The Clinical Psychologist, 37, 29-30.

Fantuzzo, J. W., & Moon, G. W. (1984). Competency mandate: A model for teaching skills in the administration of the WAIS-R. Journal of Clinical Psychology, 40, 1053-1059.

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, 224-231.

Ferguson, G. A. (1971). Statistical analysis in psychology and education. New York: McGraw-Hill Book Company.

Franklin, M. R., Stillman, P. L., Burpeau, M. Y., & Sabers, D. L. (1982). Examiner error in intelligence testing: Are you a source? Psychology in the Schools, 19, 563-569.

Hanna, G. S., Bradley, F. O., & Holen, M. C. (1981). Estimating major sources of measurement error in individual intelligences scales: Taking our heads out of the sand. Journal of School Psychology, 19, 370-376.

Hajzler, D.J. (1987). Clerical errors in the scoring of the WISC-R. The Australian Educational and Developmental Psychologist, 4, 24-27.

Kaplan, R.M., & Saccuzzo, D.P. (1989). Psychological testing: Principles, applications and issues. Pacific Grove, California: Brooks/Cole Publishing Company.

Levenson, R. L., Golden-Scaduto, C. J., Aiosa-Karpas, C. J., & Ward, A. W. (1988). Effects of examiners' education and sex on presence and type of clerical errors made on WISC-R protocols. Psychological Reports, 62, 659-664.

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

Miller, C. K., Chansky, N. M., & Gredler, G. R. (1970). Rater agreement on WISC protocols. Psychology in the Schools, 7, 190-193.

Moon, G. W., Blakey, W. A., Gorsuch, R. L., & Fantuzzo, J. W. (1991). Frequent WAIS-R adminstration errors: An ignored source of inaccurate measurement. Professional Psychology: Research and Practice, 22, 256-258.

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.

Moscato, E. M., & Cummings, J. A. (1982). A checklist for WISC-R scoring accuracy. Trainer's Forum, 1, 6.

Oakland, T., Lee, S. W., & Axelrad, K. M. (1975). Examiner differences on actual WISC protocols. Journal of School Psychology, 13, 227-233.

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

Patterson, M., Slate, J. R., Jones, C. H., & Steger, H. S. (1991). The effects of practice administrations in learning to administer and score the WAIS-R. Paper presented at the Mid-South Educational Research Association. Lexington, KY. November 14, 1991.

Piotrowski, C., & Keller, J. W. (1989). Psychological testing in outpatient mental health facilities: A national study. Professional Psychology: Research and Practice, 20, 423-425.

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

Sattler, J. M. (1990). Assessment of children (3rd edition). San Diego: Jerome M. Sattler, Publisher.

Sherrets, S., Gard, G., & Langner, H. (1979). Frequency of clerical errors on WISC protocols. Psychology in the Schools, 16, 495-496.

Slate, J. R., & Chick, D. (1989). WISC-R examiner errors: Cause for concern. Psychology in the Schools, 26, 78-83.

Slate, J. R., & Hunnicutt, L. C. (1988). Examiner errors on the Wechsler scales. Journal of Psychoeducational Assessment, 6, 280-288.

Slate, J. R., & Jones, C. H. (in press). Strategies to reduce examiner errors on the Wechsler scales. Social and Behavioral Sciences Documents.

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

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). 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. (1989). Can teaching of the WISC-R be improved? Quasi-experimental exploration. Professional Psychology: Research and Practice, 20, 408-410.

Slate, J. R., Jones, C. H., Coulter, C., & Covert, T. L. (1992). Practitioners' administration and scoring of the WISC-R: Evidence that we do err. Journal of

School Psychology, 30, 77-82.

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

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.

Wagoner, R., & Slate, J. R. (1988). WISC-R examiner errors: Cause for concern. Paper presented at the Southeastern Psychological Conference, New Orleans, LA.

Warren, S. A., & Brown, W. G. (1973). Examiner scoring errors on individual intelligence tests. Psychology in the Schools, 10, 118-121.

Wechsler, D. (1991). Wechsler Intelligence Test for

Children - Third Edition. New York: The Psychological Corporation.

Wechsler, D. (1989). Manual for the Wechsler Preschool and Primary Scale of Intelligence - Revised. New York: The Psychological Corporation.

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale-Revised. New York: The Psychological Corporation.

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

Wechsler, D. (1955). Manual for the Wechsler Adult Intelligence Scale. New York: The Psychological Corporation.

Wechsler, D. (1949). Manual for the Wechsler Intelligence Test for Children. New York: The Psychological Corporation.

APPENDIX A

LAKEHEAD UNIVERSITY

955 Oliver Road, Thunder Bay, Ontario, Canada P7B 5E1

Office of the President
Telephone (807) 343-8200

9 February 1993

Dr. T. Thompson
Department of Psychology

Dear Dr. Thompson:

Based on the recommendation of the Ethics Advisory Committee, I am pleased to agree to the extension of your research target group to include psychology professionals as outlined in your letter of Jan. 29, 1993 to Dr. Connie H. Nelson.

Thank you for keeping the Ethics Advisory Committee clearly informed about revisions to your original application.

Sincerely,

DR. R. G. ROSEHART
President

/lw

APPENDIX B

APPENDIX C

CONSENT FORM

I, ________________________________ have read and understand the introductory letter of the research study entitled "An Evaluation of Methods of Training" by Dr. Thompson and his graduate student, Christine Hodgins. I agree to participate in this study. I understand, however, that my participation is voluntary and that I can withdraw from this study at any time.

____________________________ ________________

Signature of the Participant Date

CONSENT FORM

I, ________________________________ have read and understand the introductory letter of the research study entitled "An Evaluation of the Compu-check-form" by Dr. Thompson and his graduate student, Christine Hodgins. I agree to participate in this study. I understand, however, that my participation is voluntary and that I can withdraw from this study at any time.

____________________________ ______________

Signature of the Participant Date

APPENDIX D

COMPU-CHECK-FORM

Subject Name ______________________

Checked by ________________________

AGE CALCULATION CHECK

© Remember

1 year = 12 months
1 month = 30 days

		YEAR	MONTH	DAY
SUBTRACT	DATE TESTED$_A$	____	____	____
	BIRTH DATE$_A$	____	____	____
ADD	TEST AGE	____	____	____
	BIRTH DATE$_B$	____	____	____
	DATE TESTED$_B$	____	____	____

Test Age Matches Record Form ____

RAW SCORE CHECKS & SCALED SCORE CHECKS

1: INFORMATION

Add Down

	x 0 =	
	x 1 =	
29		
Item Total		Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
-	19
29	18
-	17
28	16
27	15
26	14
25	13
23-24	12
22	11
19-21	10
17-18	9
15-16	8
13-14	7
9-12	6
6-8	5
5	4
4	3
3	2
0-2	1

INFORMATION SCALED SCORE

Scaled Score Matches Record Form______

2: PICTURE COMPLETION

Add Down

	x 0 =	
	x 1 =	
20		
Item Total		Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
-	19
-	18
20	17
-	16
-	15
19	14
-	13
18	12
17	11
16	10
15	9
14	8
13	7
11-12	6
8-10	5
5-7	4
3-4	3
2	2
0-1	1

PICTURE COMPLETION SCALED SCORE

Scaled Score Matches Record Form______

3: DIGIT SPAN

Add Down

	x 0 =	
	x 1 =	
	x 2 =	
14		
Item Total		Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
28	19
27	18
26	17
25	16
24	15
22-23	14
20-21	13
18-19	12
17	11
15-16	10
14	9
12-13	8
11	7
9-10	6
8	5
7	4
6	3
3-5	2
0-2	1

DIGIT SPAN SCALED SCORE

Scaled Score Matches Record Form______

4: PICTURE ARRANGEMENT

Add Down

	x 0 =	
	x 1 =	
	x 2 =	
10		
Item Total		Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
-	19
-	18
20	17
-	16
19	15
-	14
18	13
17	12
15-16	11
14	10
13	9
11-12	8
8-10	7
5-7	6
3-4	5
2	4
-	3
1	2
0	1

PICTURE ARRANGEMENT SCALED SCORE

Scaled Score Matches Record Form______

5: VOCABULARY

Add Down

	x 0 =	
	x 1 =	
	x 2 =	
35		

Item Total — Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
70	19
69	18
68	17
66-67	16
65	15
63-64	14
60-62	13
55-59	12
52-54	11
47-51	10
43-46	9
37-42	8
29-36	7
20-28	6
14-19	5
11-13	4
9-10	3
6-8	2
0-5	1

VOCABULARY SCALED SCORE

Scaled Score Matches Record Form______

7: ARITHMETIC

Add Down

	x 0 =	
	x 1 =	
	x 2 =	
14		

Item Total — Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
-	19
-	18
19	17
-	16
18	15
17	14
16	13
15	12
13-14	11
12	10
11	9
10	8
8-9	7
6-7	6
5	5
4	4
3	3
1-2	2
0	1

ARITHMETIC SCALED SCORE

Scaled Score Matches Record Form______

6: BLOCK DESIGN

Add Down

	x 0 =	
	x 1 =	
	x 2 =	
	x 4 =	
	x 5 =	
	x 6 =	
	x 7 =	
9		

Item Total — Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
51	19
-	18
50	17
49	16
47-48	15
44-46	14
42-43	13
38-41	12
35-37	11
31-34	10
27-30	9
23-26	8
20-22	7
14-19	6
8-13	5
3-7	4
2	3
1	2
0	1

BLOCK DESIGN SCALED SCORE

Scaled Score Matches Record Form______

8: OBJECT ASSEMBLY

Add Down

	x 0 =	
	x 1 =	
	x 2 =	
	x 3 =	
	x 4 =	
	x 5 =	
	x 6 =	
	x 7 =	
	x 8 =	
	x 9 =	
	x 10 =	
	x 11 =	
	x 12 =	
4		

Item Total — Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
-	19
41	18
-	17
40	16
39	15
38	14
37	13
35-36	12
34	11
32-33	10
30-31	9
28-29	8
24-27	7
21-23	6
16-20	5
13-15	4
9-12	3
6-8	2
0-5	1

OBJECT ASSEMBLY SCALED SCORE

Scaled Score Matches Record Form______

9: COMPREHENSION

Add Down

☐	x 0 =	◯
☐	x 1 =	◯
☐	x 2 =	◯
16		
Item Total		Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
32	19
31	18
-	17
30	16
29	15
27-28	14
26	13
25	12
23-24	11
21-22	10
19-20	9
17-18	8
14-16	7
11-13	6
8-10	5
6-7	4
4-5	3
2-3	2
0-1	1

COMPREHENSION SCALED SCORE

Scaled Score Matches Record Form______

10: DIGIT SYMBOL

Add Down

☐	x 0 =	◯
☐	x 1 =	◯
93		
Item Total		Raw Score Total

Raw Score Matches Record Form ______

Raw Score	Scaled Score
93	19
91-92	18
89-90	17
84-88	16
79-83	15
75-78	14
70-74	13
66-69	12
62-65	11
57-61	10
53-56	9
48-52	8
44-47	7
37-43	6
30-36	5
23-29	4
16-22	3
8-15	2
0-7	1

DIGIT SYMBOL SCALED SCORE

Scaled Score Matches Record Form______

11: SIMILARITIES

Add Down

☐	x 0 =	◯
☐	x 1 =	◯
☐	x 2 =	◯
14		
Item Total		Raw Score Total

Raw Score Matches Record Form: ______

Raw Score	Scaled Score
-	19
28	18
-	17
27	16
26	15
25	14
24	13
23	12
22	11
20-21	10
18-19	9
16-17	8
14-15	7
11-13	6
7-10	5
5-6	4
2-4	3
1	2
0	1

SIMILARITIES SCALED SCORE

Scaled Score Matches Record Form______

VERBAL SCORE CALCULATION CHECK

		Add Down
1.	Information	⬭
3.	Digit Span	⬭
5.	Vocabulary	⬭
7.	Arithmetic	⬭
9.	Comprehension	⬭
11.	Similarities	⬭
	VERBAL SCORE	

VERBAL SCORE

⬡

Verbal Score Matches Record Form________

'ERFORMANCE SCORE CALCULATION CHECK

		Add Down
2.	Picture Completion	⬭
4.	Picture Arrangement	⬭
6.	Block Design	⬭
8.	Object Assembly	⬭
10.	Digit Symbol	⬭
	PERFORMANCE SCORE	

PERFORMANCE SCORE

⬡

Performance Score Matches Record Form________

APPENDIX E

Your Opinion of the Compu-check-form

When scoring WAIS-R protocols, do you typically check addition, raw to scale score conversion, etc.?

yes ___
no ___ (proceed to question # 3)

When checking WAIS-R protocols do you simply repeat the scoring and score conversion procedures or do you use alternative (different) methods of checking?

repeat procedures ___
alternative methods ___

In your opinion do you see the compu-check-form as having any value? Please elaborate on your opinion.

4. Would you personally use the compu-check-form in your practice?

yes ___
no ___

Would you recommend that psychology students under your supervision use the compu-check-form when scoring WAIS-R protocols?

yes ___
no ___

Would you personally use some of the compu-check-form checking procedures without actually using the compu-check-form?

yes ___ (proceed to question # 7)
no ___ (proceed to question # 8)

Which compu-check-form techniques would you use?

8. Any additional comments you might have or suggestions for improving the compu-check-form are welcomed.