DOCUMENT RESUME

ED 296 263   CG 020 955

AUTHOR   Slate, John R.
TITLE   Teaching the WISC-R: An Effective Instructional
                Design Procedure.
PUB DATE   86
NOTE   27p.
PUB TYPE   Reports - Research/Technical (143)

EDRS PRICE   MF01/PC02 Plus Postage.
DESCRIPTORS   Clinical Psychology; *Error of Measurement; *Graduate
                Students; Higher Education; *Intelligence Tests;
                *Scoring; *Teaching Methods; *Testing Problems
IDENTIFIERS   *Wechsler Intelligence Scale for Children
                (Revised)

ABSTRACT
        Studies have revealed significant problems in
correctly scoring ambiguous verbal responses to test items on the
Wechsler Intelligence Scale for Children-Revised (WISC-R). This study
evaluated the effectiveness of an instructional design procedure
developed to reduce examiner scoring errors on the WISC-R. Data
concerning frequent sources of error on the WISC-R were obtained from
14 clinical psychology graduate students (pre-intervention group)
enrolled in an Individual Intelligence Testing course. Remedial
strategies were developed to clarify response categories and minimize
errors due to carelessness. Strategies were used with 9
post-intervention students who took the course the following fall
term. Pre- and post-intervention students completed the course and
administered the WISC-R several times. A total of 98 pre-intervention
and 63 post-intervention protocols were analyzed to determine the
number and type of errors. Compared to the pre-intervention group,
the post-intervention group had significantly fewer errors, with
careless mathematical and clerical mistakes almost eliminated.
Students continued to make mistakes, showing no improvement with
practice in test administrations. A certain amount of difficulty in
scoring may be inherent in the WISC-R test manual, and it may be
advisable to incorporate examiner scoring errors into the WISC-R
standard error of measurement. (Author/NB)

ERIC
Full Text Provided by ERIC

Teaching the WISC-R: An Effective

Instructional Design Procedure

John R. Slate

Arkansas State University

Running Head: WISC-R INSTRUCTION

2

Abstract

The effectiveness of an instructional design procedure developed
to reduce examiner scoring errors on the WISC-R was
investigated. Scoring errors were significantly reduced, with
careless mathematical and clerical mistakes almost eliminated.
Corrected Full Scale IQ scores were almost all within +/- 2
points of the originally assigned IQ scores. Even so, subjects
continued to make errors, showing no improvement with practice
in test administrations. It may be that a certain amount of
difficulty in scoring is inherent in the WISC-R test manual
(e.g., correctly scoring ambiguous responses). Enough data
exists to support the notion that errors are commonplace;
perhaps, then, we need to incorporate examiner scoring errors
into the WISC-R standard error of measurement.

Teaching the WISC-R: An Effective

Instructional Design Procedure

Numerous studies (e.g., Bradley, Hanna, & Lucas, 1980;

Brannigan, 1975; Conner & Woodall, 1983; Franklin, Stillman,

Burpeau, & Sabers, 1982; Miller & Chansky, 1972) have shown

significant problems exist in correctly scoring ambiguous verbal

responses to test items on the Wechsler Intelligence Scale for

Children-Revised (WISC-R). Moreover, clerical and mathematical

errors are not unusual, even on professional psychologists' test

protocols (e.g., Oakland, Lee, & Axelrad, 1975; Sherrets, Gard,

& Langner, 1979). One primary cause of examiner scoring error

may be inadequate training and poor instructional design

procedures of intelligence testing courses. Relatively little

has been done in the area of training students to score the

WISC-R more accurately (Blakey, Fantuzzo, Gorsuch, & Moon,

1987); what has been conducted has been reported as ineffective

(Conner & Woodall, 1983; Warren & Brown, 1973). Poor

instructional preparation in intelligence testing courses may

produce examiners who do not have sufficient knowledge of test

manual instructions nor an awareness of the significance of

standardized procedures.

Research studies (Dana, Gilliam, & Dana, 1976; Levitt,
1973; Rice & Gurman, 1973; Russ, 1978; Shemberg & Keeley, 1974;
Sturgis, Verstegen, Randolph, & Garvin, 1980) exist which
support the hypothesis that training appears to be less than
adequate. For example, Garfield and Kurtz (1973) reported that
the primary skill deficit of clinical graduate students was a
lack of assessment skills. Though competently administering
standardized tests is a basic assessment skill, surveys of
internship directors (e.g., Dana et al., 1976; Shemberg &
Keeley, 1974; Sturgis et al., 1980) cited assessment skills as
among new interns' most prominent skill deficits. Two specific
features of training mentioned as deficient are (a) inadequate
teaching and (b) disparaging attitude toward diagnostic testing.
Evidence that inadequate teaching may be an important variable
in poor assessment skills comes from several sources (Dana et
al., 1976; Drabman, 1985; Sturgis et al., 1980). According to
Drabman (1985), students "often arrive at their internship sites
not knowing how to administer, score, and interpret" commonly
administered tests (p. 624). In regard to (b), Garfield and
Kurtz (1973) commented that "university training tends to make
students have an overly critical attitude toward diagnostic
testing" (p. 352). Thus, students may leave assessment courses

5

not only with minimal testing skills but also with a distaste

for testing which may further impair accurate administration and

scoring of intelligence test protocols.

Even in the studies (e.g., Boehm, Duker, Haesloop, & White,

1974; Fantuzzo, Sizemore, & Spradlin, 1983) that examined

instructional design procedures in intelligence testing,

graduate students were still making errors following course

completion; errors which later may become more serious due to

less supervision and more demands for the practitioner's limited

time. The conclusion cannot be made that students will become

more proficient in test administration and scoring following

graduation, particularly if they lack competency prior to

graduation. The suggestion (Franklin, Stillman, Burpeau, &

Sabers, 1982) that present training methods need to be

re-examined is still apt today. The purpose of the present

research was to examine the effectiveness of an instructional

design procedure developed to minimize frequently occurring

scoring errors on the Wechsler Intelligence Scale for

Children-Revised.

6

Method

Subjects

Two groups of graduate students enrolled in a clinical psychology master's program at a southeastern university and in an Individual Intelligence Testing course served as subjects. Fourteen subjects were enrolled in the 1985 fall semester course (pre-intervention group) while 9 subjects were enrolled in the 1986 fall semester course (post-intervention). Four were male and 10 were female in the 1985 group, while 3 were male and 6 were female in the 1986 group. Academic scores (e.g., Graduate Record Examination and college grade point averages) were commensurate for both groups.

Procedure

Data were obtained from the first group of 14 subjects concerning frequent sources of error on the WISC-R. Following the determination of subtest items that were most difficult to score and reasons errors occurred on those items, remedial strategies were developed. These strategies were aimed at clarifying response categories (e.g., point value assignments) as well as minimizing error due to carelessness. These remedial strategies were used in the following fall to ascertain their effectiveness in decreasing examiner scoring errors.

For both groups, students were required to study the test
manual prior to observing a practice demonstration of the
WISC-R. Problems in administration and errors in scoring that
might occur were discussed. For the post-intervention group,
the remedial strategies were examined point by point and each
student was provided with a written copy. All subjects for both
groups were required to administer the WISC-R 8 times to child
and/or adolescent volunteers. The average number of WISC-R
protocols completed in assessment courses is 7.3 (Oakland &
Zimmerman, 1986). Students were paired together with each one
responsible for checking the other's protocols for errors prior
to submission to the instructor for grading. The form utilized
by the students in evaluating WISC-R protocols was a
modification of one used in a study by Conner and Woodall
(1983)[1]. Both written and verbal feedback were reported to
students by the instructor following each of the 7 WISC-R
protocols so that corrections could be made prior to the final
administration observed by the instructor.

Following completion of the class, protocols were analyzed
to determine the number and type of errors made by the graduate
students. A total of 98 and 63 protocols were analyzed for the
pre- and post-intervention groups, respectively.

Results

A t-test performed on the mean errors per test administration ($t=8.36$, $p<.01$) found a significant difference in mean errors for the pre- and post-intervention groups. When examining the means of scoring errors by test administration (Table 1), one notices that the post-intervention group mean is significantly lower on the first test administration. This decrease was maintained throughout the remaining six administrations. Next, a repeated measures ANOVA was used to analyze whether errors significantly decreased over time for the pre- and post-intervention groups. F's of 1.21 ($p<.31$) and 1.01 ($p<.43$) were obtained for the pre- and post-intervention groups, respectively, indicating that scoring errors did not decrease over test administrations. An examination of the means reveals that the intervention provided an immediate reduction in scoring errors ($M=3.88$ for the post-intervention group compared to $M=9.57$ for the pre-intervention group), but did not result in further decreases over time. This finding has implications for instructional design procedures for assessment courses.

_____

Insert Table 1 about here

_____

The mean scoring error for each Wechsler subt? , for both
pre- and post-intervention groups is shown in Table 2. As found
in previous research (e.g., Miller & Chansky, 1972), Vocabulary,
Comprehension, and Similarities were the three subtests in which
students made the most mistakes. The post-intervention groups'
mean errors per subtest was about half or less of the errors
made by the pre-intervention group. The maximum number of
scoring errors made on any one protocol decreased from 33 in the
pre-intervention group to 9 in the post-intervention group.

Insert Table 2 about here

In the pre-intervention group, errors on 32.6% of test
protocols did not influence the Full Scale IQ score while the IQ
scores on 58.7% of the protocols were originally assigned scores
1 to 5 points higher than the corrected IQ score. Prior to the
intervention, students were more likely to assign too many
points to an examinee's answer rather than too few points. In
the post-intervention group, 68.3% of the protocols had no
change in the Full Scale IQ score. The remaining IQ scores were
almost all (exception of 3.4%) within 2 points of the corrected
IQ score.

The most frequent types of mistakes were examined and
ranked in Table 3. Inappropriate questioning and assigning too
many points for an examinee's response are the two most
frequently occurring errors for both pre- and post-intervention
groups. Several mechanical and clerical errors such as no red
pencil for coding, incorrect subtest total, and failur: to
record examinee responses were noticeably reduced in the
post-intervention group.

_____

Insert Table 3 about here

_____

Tables 4 and 5 indicate the most frequent incorrectly
scored subtest items and subtests by categories of errors and
suggestions designed to minimize those errors. On Table 5,
numerous items and subtests were scored consistently wrong for
the pre-intervention group. Following the remedial strategies,
the post-intervention group shows a reduction in the number of
difficult items and number of subtests with consistent error
patterns. Samples of the remedial strategies are provided in
Table 52.

_____

Insert Tables 4 and 5 abcut here

_____

Discussion

Directing education toward likely sources of error on the
WISC-R appeared to be an effective procedure for decreasing
examiner scoring error. Making students aware of the existence
of errors, common difficulties, and reasons for those
difficulties seemed to have an immediate and lasting effect.
Errors were cut in half on the first test administration and
remained so for the next six administrations. Even with the
strategies, students averaged between 3 and 4 errors per
protocol. Some of these errors may be caused by ambiguity in
the test manual and may be difficult to modify. Brannigan
(1975) ard Miller and Chansky (1972) suggested a revision of
Wechsler test items most subject to ambiguous replies. Even the
Revised version appears in need of clarification because many
verbal responses given by children are not scorable clearly by
the test manual and examiners "read into" the responses.

Students did not significantly reduce their scoring errors
over time. That is, test administrations alone did not result
in fewer scoring errors. This was true for both pre- and

post-intervention groups. The adage of "practice makes perfect" does not seem to accurately reflect the acquisition of competent assessment skills. What may be happening is that students learn "bad habits" and, therefore, continue to make the same mistakes time and time again. Thus, we should not conclude that students will become more proficient with practice in scoring WISC-R test protocols when the research evidence indicates that this is not the case (e.g., Bradley et al., 1980; Oakland et al., 1975; Sherrets et al., 1979).

One limitation of the present study involves the composition of the sample. A limited number of subjects were used in the study. Also, all the subjects herein were enrolled in a master's level clinical psychology program. Thus, the results may not be generalizable to other graduate students or to other psychological specialty areas. To determine the applicability of these findings, research needs to be conducted with other student groups such as school psychology and with larger samples. Moreover, it is possible that the subjects in the post-intervention group were different from those in the pre-intervention group in ways related to assessment skills. That is, the post-intervention group might have made fewer scoring errors even without the remedial strategies, simply

13

because of unique subject variables. Therefore, this study
needs to be replicated to more clearly ascertain its
generalizability.

In summary, improved skills in psychological assessment is
a primary need for professional development (Anderson, Cancelli,
& Kratochwill, 1984). Strategies designed to reduce examiner
scoring errors on WISC-R protocols appeared to be effective.
Students made fewer mistakes, resulting in IQ scores that were
almost all within +/- 2 points of the corrected IQ score. Even
with the instructional design method discussed herein, students
continued to make errors which practice did not decrease. This
finding may reflect ambiguity in the WISC-R test manual rather
than poor educational procedures. Given the research base to
date, it is time to consider incorporating examiner scoring
error into the WISC-R's standard error of measurement. To fail
to do otherwise ignores reality.

References

Anderson, T., Cancelli, A., & Kratochwill, T. (1984).
Self-reported assessment practices of school psychologists:
Implications for training and practice. Journal of School
Psychology, 22, 17-29.

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

Boehm, A., Duker, J., Haesloop, M., & White, M. (1974).
Behavioral objectives in training for competence in the
administration of individual intelligence tests. Journal of
School Psychology, 12, 150-157.

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

Brannigan, G. (1975). Scoring difficulties on the Wechsler
Intelligence Scale. Psychology in the Schools, 12, 313-314.

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

Dana, R., Gilliam, M., & Dana, J. (1976). Adequacy of academic

   clinical preparation for internships. Professional

   Psychology, 7, 112-116.

Drabman, R. (1985). Graduate training of scientist-practitioner

   -oriented clinical psychologists: Where we can improve.

   Professional Psychology: Research and Practice, 16, 623-633.

Fantuzzo, J., Sizemore, T., & Spradlin, W. (1983). A

   competency-based model for teaching skills in the

   administration of intelligence tests. Professional

   Psychology, 14, 224-231.

Franklin, M., Stillman, P., Burpeau, M., & Sabers, D. (1982).

   Examiner error in intelligence testing: Are you a source?

   Psychology in the Schools, 19, 563-569.

Garfield, S., & Kurtz, R. (1973). Attitudes toward testing: A

   survey of directors of internship training. Journal of

   Consulting and Clinical Psychology, 40, 350-355.

Levitt, E. (1973). Internship versus campus: 1964 and 1971.

   Professional Psychology, 4, 129-132.

Miller, C., & Chansky, N. (1972). Psychologists' scoring of

   WISC protocols. Psychology in the Schools, 9, 144-152.

Oakland, T., Lee, S., & Axelrad, K. (1975). Examiner

differences on actual WISC protocols. Journal of School

Psychology, 13, 227-233.

Oakland, T., & Zimmerman, S. (1986). The course on individual

mental assessment: A national survey of course instructors.

Professional School Psychology, 1, 51-59.

Rice, D., & Gurman, A. (1973). Unresolved issues in the

clinical psychology internship. Professional Psychology, 4,

403-408.

Russ, S. (1978). Teaching psychological assessment: Training

issues and teaching approaches. Journal of Personality

Assessment, 42, 452-456.

Shemberg, K., & Keeley, S. (1974). Training practices and

satisfaction with preinternship preparation. Professional

Psychology, 5, 98-105.

Sherrets, S., Gard, G., & Langner, H. (1979). Frequency of

clerical errors on WISC protocols. Psychology in the

Schools, 16, 495-496.

Sturgis, D., Verstegen, J., Randolph, D., & Garvin, R. (1980).

Professional psychology internships. Professional

Psychology, 11, 567-573.

Warren, S., & Brown, W. (1972). Examiner scoring errors on
   individual intelligence tests. _Psychology in the Schools_, _9_,
   118-122.

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

Table 1

Mean Error Scores by Test Administration for Pre- and
Post-Intervention Groups.

Test Administration	Pre-	Post-
	M	M
1	9.57	3.88
2	8.62	4.38
3	9.62	3.11
4	7.54	4.12
5	7.23	1.78
6	6.85	3.78
7	7.15	3.11

Table 2

Comparison of Means, Ranges and Ranks of Examiner Errors by
Subtest per Protocol.

Subtest	Pre-			Post-		
	M	Range	Rank	M	Range	Rank
Information	.38	0-4	7	.12	0-1	7
Similarities	1.14	0-5	3	.42	0-3	3
Arithmetic	.12	0-3	10	.05	0-1	10
Vocabulary	2.22	0-11	1	1.27	(-3	1
Comprehension	1.98	0-9	2	.83	0-5	2
Picture Completion	.28	0-4	8	.20	0-3	5
Picture Arrangement	.43	0-3	6	.10	0-2	8
Block Design	.26	0-5	9	.07	0-1	9
Object Assembly	.55	0-4	4	.22	0-4	4
Coding	.54	0-3	5	.13	0-1	6
Independent Errors	8.10	0-33		3.40	0-9	
Total Errors	15.18	0-45		6.45	0-20	

Table 3

Comparisons of Error Type Rankings Across Subtests.

	Pre-		Post-	
Error Type	%	Rank	%	Rank
0 point credit for a 2/1 point answer	6.8	6	11.8	4
1 point credit for a 2/0 point answer	12.9	4	9.3	5
2 point credit for a 1/0 point answer	24.7	1	14.2	2
Inappropriate questioning	23.8	2	36.8	1
Failure to record examinee's response	15.6	3	6.4	6
Incorrect basal and/or ceiling	7.1	5	12.7	3
Incorrect credit for items below basal and/or above ceiling	1.0	9	0.5	8
Incorrect total for subtest	4.3	7	1.5	7
No red pencil for Coding	3.6	8	0.0	9

Table 4

Comparisons of Most Frequent Incorrectly Scored Subtest Items

and Subtests by Error Type.

	Pre-	Post-
Incorrect point assignment		
Information	13, 25, 26	-------
Similarities	5, 9, 14, 16	9, 16
Vocabulary	5, 7, 8, 10, 12, 14, 20	5, 7, 12
Comprehension	3-4, 6-9, 11-12, 15-16	8-9, 12, 16
Inappropriate questioning		
Information	13, 26	-------
Similarities	6	-------
Vocabulary	5, 7, 8, 10, 12, 14, 20	5, 12
Comprehension	3, 4, 7, 8, 9, 12, 16	8-9, 16
Basal and/or ceiling problems		
Arithmetic		-------
Vocabulary		
Picture Completion		-------
Picture Arrangement		-------
Block Design		-------

(Table Continues)

Table 4

Incorrect subtest total

  Object Assembly   --------

  Coding

Note. -------- means that no consistent pattern of errors was found.

Table 5

Sample Comments from the WISC-R Suggestions for Remediation.

Subtest Item	Comment
Information	
13	Most errors were assigning 0 points for a 1 point response. An examinee response of the stomach performing some activity on food receives 1 point.
Similarities	
5	All errors were assigning 2 points for 1 point answers. To earn 2 points, the verbal comment must indicate that both are fruits. One point answers indicate specific properties, uses, or other general classifications.
14	Majority of mistakes were assigning 0 points for a 1 point response. Two point answers have to indicate abstract concepts or social ideas while 1 point answers show civil rights, have to do with freedom, democracy, or symbols.

(Table Continues)

Table 5

Subtest Item	Comment
Vocabulary	
5	Most errors were assigning 2 points for a 1 point answer. For 2 points, the child must indicate the object's general conceptualization or two 1 point responses. All 1 point answers in the manual are (Q).
7	The meaning of alphabet is letters in a language that are used in words, to write, and have sounds. A 1 point reply concerns letters or ABC's and are (Q). Note that if the child recites part or all of the alphabet, s/he receives 1 point.
Vocabulary	
8	Two points are assigned when the examinee states a donkey is like a horse but different in some way or indicates its general classification as an animal. One point is given when the child states a use or describes specific attributes.

(Table Continues)

Table 5

Subtest Item	Comment
Comprehension

3   Examiners made mistakes in failing to (Q) when necessary and assigning 2 poi..ts for a 1 point answer. To earn 2 points, it must indicate both general areas on page 177. A reply of 1 area earns 1 point.

11   Examinee must indicate a correct general statement that suggests awareness of the significance of meat inspectors for the public to earn 2 points. One point comments concern a specific statement that points cut advantages of having or the dangers of not having meat inspectors but lack implications for society at large.

Coding

General   Make sure you have several pencils with red lead for the examinee. Do not score from memory; use the scoring stencil. If it is not available, use the key and check each item one at a time.

Footnote

[1] A copy of the modified Conner and Woodall (1983) form is available to interested readers. Please address inquiries to the author.

[2] A copy of the remedial strategies is available to interested readers. Please address inquiries to the author.