



Appendix B

Studies of Technical Quality





Studies of Technical Quality

A series of empirical studies were carried out throughout a seven year period to validate the IOWAN. The approach followed the approach used in the late 1970s and early 1980s at the University of Minnesota Institute for Research on Learning Disabilities for developing and validating Curriculum-Based Measurement (CBM) procedures in basic skills areas. The overarching question guiding the study was:

Can a standard indicator be developed that would be useful for screening and progress monitoring purposes in the area of handwriting?

A series of characteristics were set forth *a priori* that any potential measure had to meet in order to be useful for these purposes. It needed to be:

- Brief to administer, so that it could be used frequently
- Repeatable, which required that it have many alternate test forms
- Easy to create, to increase the usability and feasibility of measurement
- Easy to score, to facilitate its use by persons with minimal training
- Standardized, that is, administered and scored in the same way each time
- Technically adequate, that is, demonstrate reliability, validity and fairness in relation to the intended purposes of screening and progress monitoring.

At the outset of this project, a review of available assessments in the area of handwriting was conducted to determine the stimulus characteristics and response demands of tasks typically used to measure handwriting. Based on this review, a near point copying task was selected as the focus for this project. Other possibilities examined included scoring students' existing classroom work, asking students to write the alphabet and asking students to write numbers. Near point copying was selected initially because stimulus materials could be controlled, the level of materials could be varied and made appropriate for various grades, and this task did not rely on prior knowledge to complete. While the task is not a pure motor task in that there is some cognitive mediation involved (the individual needs to see the letters and translate this visual cue into an appropriate motor response), this task was as close to pure motor task as could be identified without requiring the individual to construct a written product in response to no stimulus.

Stimulus Characteristics and Response Demands: The IOWAN can be administered either in an individual or group setting. The IOWAN task requires the individual to sit at a table, hold an appropriate writing utensil (of the individual's choosing) and look at a typewritten story – one that is grade-level appropriate and one that is a common passage administered to all respondents. These passages are administered at the same sitting in two successive administrations. In addition to the typewritten story, the individual is provided a piece of writing paper that is typical for her/his handwriting work. The individual is read a standardized set of instructions and asked to copy as much of the story on to the blank page as he/she can in two minutes. At the end of two minutes, the individual is asked to stop and put her/his pencil down.

Validation Studies

A series of studies were conducted over a number of years to examine the technical characteristics of the measures. Both feasibility and psychometric issues were examined. This document contains a brief description of each study conducted.

Study 1: Pilot of Materials and Administration/Scoring Procedures

Purpose: The purpose of this pilot was to generate measurement materials, administration and scoring procedures that would be used in subsequent evaluations.



Method: A group of professionals with expertise in education, occupational therapy, and research/evaluation met to select an initial group of passages and to construct administration and scoring procedures. Initial materials were selected from a set of progress monitoring passages previously published by Stan Deno and Doug Marston from the University of Minnesota. Permission for using the passages as part of the IOWAN was obtained. Standard CBM administration and scoring procedures described by Shinn (1989) were adapted for use in the IOWAN project. Once the initial materials and procedures were created, five occupational therapists administered the materials to students ranging from grade 1 to grade 5 to examine the feasibility, understandability and usability of the procedures.

Results: Student performance data from this study were not examined. Results indicated that administration of the IOWAN probes took roughly 5 minutes per child when transition time was included in the time estimate. Also, the instructions were deemed understandable by all of the children in the study. Some of the first grade students needed additional paraphrasing of the instructions, which broke standardization, but was not thought to invalidate the procedure. Initial scoring procedures were examined as well. The workgroup came together to pilot the initial scoring criteria with the responses generated from this first group. A number of changes in scoring procedures were identified by this initial pilot group and these changes were worked into the initial administration and scoring trainings.

Discussion: By the end of this initial pilot, it appeared that the initial pool of stimulus passages were appropriate for additional study. The administration procedures worked well and the scoring procedures, though in need of additional refinement, appeared consistent enough for additional data to be collected. Because scoring was done on permanent products, if changes were needed in scoring procedures in the future, earlier student writing samples could be rescored retrospectively to rerun earlier analyses to determine if scoring changes had any effect on overall technical adequacy of the IOWAN.

Study 2: Initial Progress Monitoring Data Collection

Purpose: The purpose of this study was to collect multiple samples of progress monitoring data from different practitioners and different students across Iowa to provide initial data for technical adequacy investigations. Specifically, this study examined the reliability and validity of the near point copying task. It also generated a large sample of time series that could be examined directly and their measurement characteristics measured directly (e.g., level, slope and variability)

Method: Occupational Therapists and Occupational Therapy Assistants (OTA) from each of Iowa's 15 AEAs were invited to participate in this study. Each occupational therapist or occupational therapy assistant was asked to monitor the progress of two children from one grade level (all students for this study came from either second or fourth grade). One of the children monitored was to be a student who was a typical hand writer, the other was a child from the same grade level who received special education services and who had a significant problem with handwriting. All participating Occupational Therapists and OTAs were provided training on the administration of the IOWAN measures. Each of these persons scored a criterion student handwriting sample at the end of their training and their level of accuracy (agreement with the same performance scored by the developers) was examined. Prior to participating in data collection with students, each participants' accuracy needed to be .90 or above. If their accuracy was below this level, additional training was provided until their scoring accuracy improved to at above the criterion. Students' progress was monitored between December and May of one school year. Data were collected one time per week on each student's performance. Three research questions were addressed:

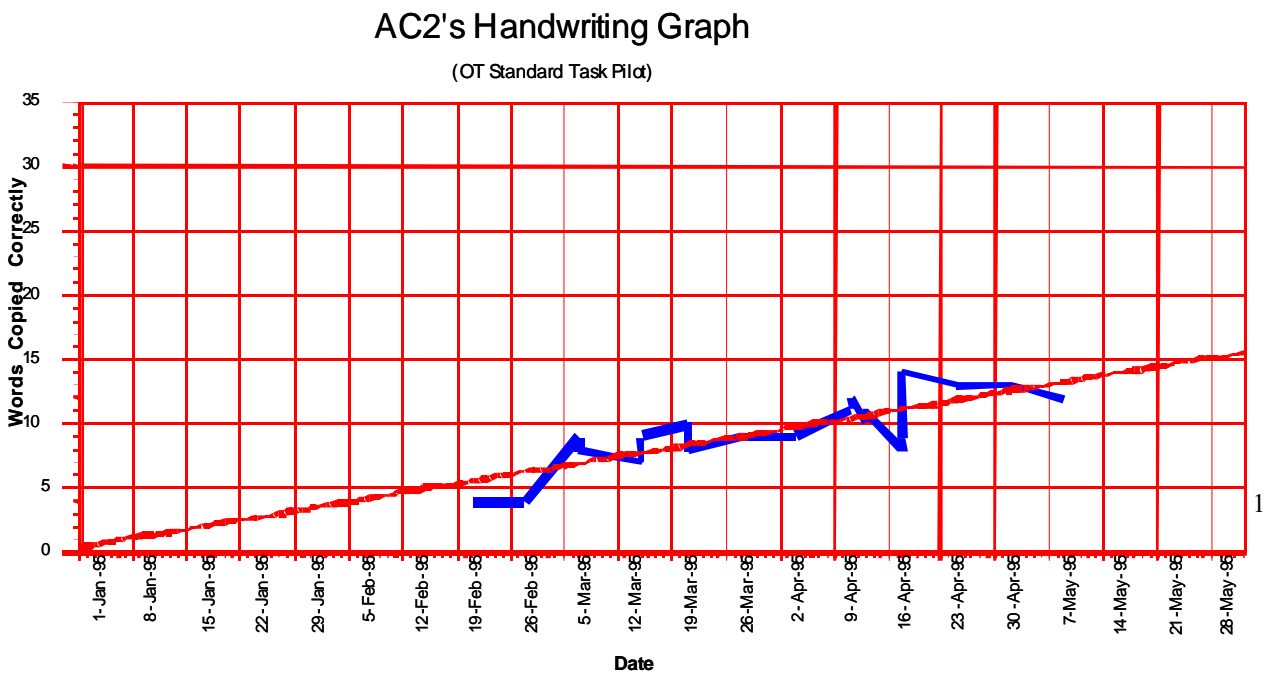
- a. Does the near-point copying task generate both within-student and between-student variability in both level and slope?
- b. Does student performance on the near-point copying task reflect known student performance differences such as those between special and general education students?
- c. Does student performance on the near-point copying task reflect developmental differences in writing proficiency by gender?



Results: Cases were considered complete if data were collected for at least 15 weeks within the data collection window. One-hundred forty-four second grader's data and 77 fourth grader's data were returned to the researchers. A series of analyses were conducted and are listed below.

1. Research Question a: Individual students rates of progress were estimated using typical Curriculum-Based Measurement formative evaluation procedures. Individuals who had 8 or more data points collected throughout the study period were included in the analysis (144 second grade students' data met this criterion, 77 fourth grade students met this criterion). The range of slopes within each grade level and levels of performance within each grade level were examined. Slopes were calculated using ordinary least squares for each student and level was calculated by taking the arithmetic average of all data points in a students time series. Student slopes ranges from -1.5 words per week to 8.9 words per week. Examples of actual student performance graphs are presented in Figures 1 and 2. The average slope of improvement was estimated by averaging the slopes of performance within each grade. The average slope for 2nd graders was 0.26 words per week growth. The average slope for 4th graders was 0.28 words per week.

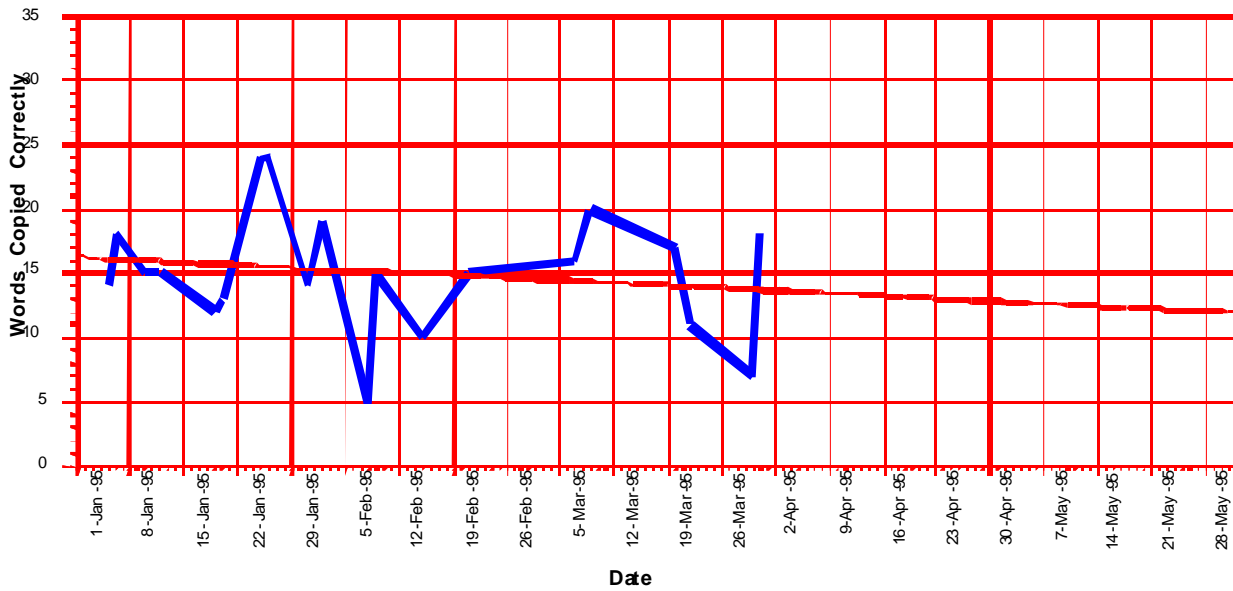
Figure 1 and 2. Example student performance graphs on the IOWAN.





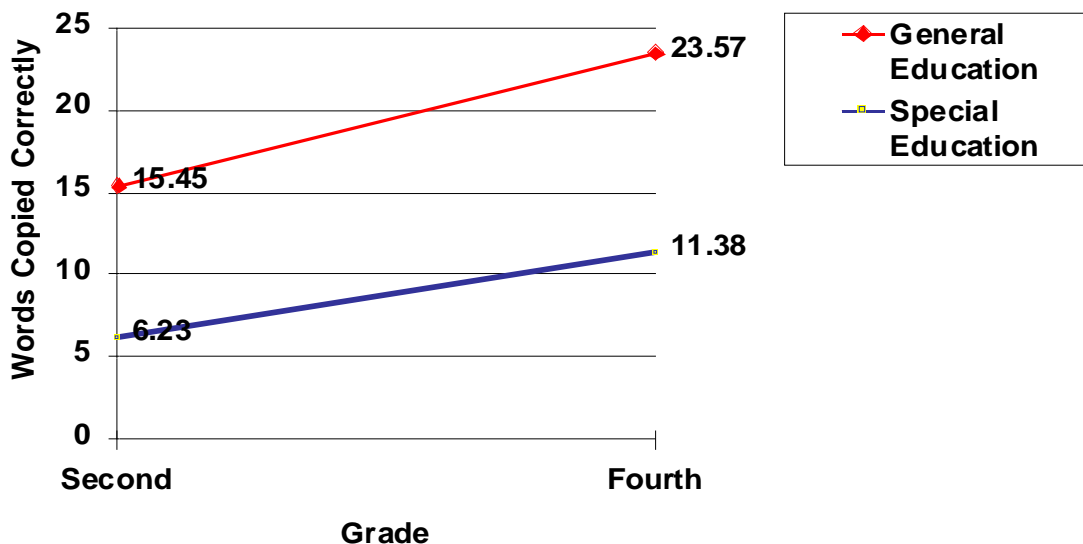
Austin's Handwriting Graph

(OT Standard Task P1d)



- Research question b: The second set of analyses examined differences in student performance by grade level (2nd and 4th) and placement status (general education, special education). A two-way analysis of variance was run using level of performance as the dependent measure. Significant differences ($p < .05$) were found for both main effects (grade and placement). The grade-by-placement interaction was not statistically significant. These findings mean that 4th graders overall performed more fluently on the near-point copying task than did second graders. It also means that students in general education also performed better than students in special education. A graphic depiction of this finding is presented in figure 3.

Figure 3: General and special education students' performance on the IOWAN.





A second analysis was performed between special and general-education student. In this analysis, slope of student performance was the dependent measure and educational status (general- vs. special-education) was the independent variable. A one way analysis of variance was conducted and the results approached statistical significance ($p < .07$). Results from this analysis are presented in Table 1. This finding suggests that students in general education may be making more progress in this sample than students placed in special education.

Table 1: Slopes of student performance differences by educational status.

Educational Status	Mean of Slopes	Number of Students
GE	0.399	117
SE	0.189	104

*Note: Slope coefficients reflect slope per week.

3. Research question c: The third set of analyses examined differences in performance as a function of gender. Both second and fourth grade students data were used in this analysis. A one-way analysis of variance was conducted using level of performance as the dependent measure and gender as the independent variable. Results indicated that girls on average correctly copied more words $p < .05$. Girls overall copied 15.29 words per assessment while boys copied 13.02.

Discussion: Results from studies 2a suggest that the near point copying task generates a broad range of slopes of student performance growth. This finding was important since it is known that student handwriting proficiency varies within the elementary school population and any valid measure of handwriting would be expected reflect similar variability in student performance. Results from study 2b suggest that general education students perform better on a near-point copying task than do special education students with school-identified handwriting difficulties. This finding was important because it documented that students with known handwriting difficulties did significantly more poorly on the IOWAN than did students who were typical handwriters. If the IOWAN is a valid measure of handwriting performance, then it would be expected to reflect known differences in student performance. Results from study 2c suggest that, on average, for girls in 2nd and 4th grade, girls near-point copying skills are slightly better than boys of the same grade-levels.

Study 3: Examination of the Reliability of the IOWAN

Purpose: The purpose of this study was to document the score reliability of the IOWAN. A series of reliability analyses were conducted. The accuracy, overall interater agreement and test-retest reliabilities were examined.

Method and Results:

Accuracy. For study 3a, accuracy is defined as the degree to which a group of scorer’s scoring results are consistent with and reflect the “correct way” that they should have scored a set of IOWAN tests. In this case, the objective standard for measuring scoring accuracy was created by the creators of the IOWAN tasks. Responses from four different students (grades 2 and 4) to four different IOWAN passages were selected from previous studies. IOWAN creators collectively scored each of these four different student responses based on the IOWAN scoring guidelines. In essence, the way that these student samples were scored represents the “correct scoring” for each of the four passages. Next, the IOWAN creators brought together a group of 86 Occupational Therapists and Occupational Therapy Assistants who were trained in administration and scoring of the IOWAN. During the training, questions were answered and many practice scoring samples were completed. This training took approximately an hour. At the end of the training, each of the 86 scorer trainees were asked to score each of the 4 passages previously scored by the IOWAN creators. Accuracy was calculated using a point-by-point agreement method. The overall accuracy rate was calculated across all 104 scorers for each passage. Accuracy coefficients were .74, .89, .79 and .97 respectively for passages 1 through 4 respectively, with a median accuracy coefficient of .85.



Overall Interrater Agreement. In study 3b, overall agreement among the same group of scorers was calculated using the overall percent agreement statistic. Results from this analysis were very similar to the accuracy statistics. The overall agreement coefficients were .71, .90, 79 and .94 overall, with a median overall interrater agreement coefficient of .84.

Test-Retest Reliability. In study 3c, 70 second graders and 54 fourth graders progress were monitored systematically, one time per week, for a minimum of 12 weeks during one school year. Progress monitoring materials were set up in a consistent order that was used with each student within each grade level. That is, one set of grade-level appropriate passages was used with the second grade students and a different set of grade-level appropriate passages were used with the fourth graders. In each set of passages, one passage appeared three times, one time during week 1, one time during week 3 and one time during week 5. Again, the passage that was repeated was grade appropriate at second and fourth grade.

Using this methodology, test-retest reliability was able to be tested multiple times for each OT/student combination. An ordered pair of scores was created for each OT using their student's scores from week 1 to week 3, creating a three week test-retest window. This process was repeated between weeks 1 and 5 creating a 5-week test retest interval. The Pearsons product-moment correlation coefficient was calculated for both the 3 and 5 week retest intervals. The three week test-retest correlation was .97 and the five week test-retest correlation was .90.

Discussion: Studies 3a, 3b and 3c demonstrated that the IOWAN scores can be collected with reasonable reliability and that acceptable levels of error are present in IOWAN scores as a function of rater and time.

Study 4: IOWAN Concurrent Validity Studies

Purpose: The purpose of study 4 was to examine the degree to which IOWAN scores were related to other measures of handwriting. If the IOWAN is a valid measure of handwriting, then scores earned on the IOWAN should be related to scores earned on other valid measures of handwriting.

Method and Results: To examine this question, the IOWAN was administered to a group of students along with two other criterion measures of Handwriting. Seventy six 2nd grade students and twenty seven 4th graders of all handwriting skill levels participated in this study. All data on each student were collected during one three week period.

Data from three measures were collected on each student. First, the student's performance on an IOWAN near-point copying task was collected. Second the student was administered the Children's Handwriting Evaluation Scale (CHES). The CHES is a standardized test of handwriting that yields both a quantitative and qualitative standard score of the student's handwriting. The quantitative standard score is related the student's rate of production, the qualitative score is related to letter formation and spacing. Third, both the student's teacher and the occupational therapist collecting the student's IOWAN data completed a survey, rating critical characteristics of the student's handwriting. The survey consisted of 12 items asking the parent and OT to rate the student's handwriting on the following characteristics: legibility, neatness, letter formation, speed, spacing, and overall skill.

The correlation between student's scores on the IOWAN and the CHES was documented using a Pearson product-moment correlation. To prevent artificial inflation due to an extension in range, second grade and fourth grade data were analyzed separately. The concurrent validity coefficients between the rate standard score on the CHES and student performance on the IOWAN are presented in Table 2.



Table 2: Concurrent validity correlation coefficients between the IOWAN and the CHES.

Grade	IOWAN and Rate Standard Score from the CHES	IOWAN and Qualitative Standard Score from the CHES
Second	.66	.86
Fourth	.72	.33

The second criterion measure, parent and teacher rating data were analyzed in a similar way. The concurrent validity coefficients between the teacher and OT ratings are presented in Table 3.

Table 3: Concurrent validity correlation coefficients between the IOWAN and Teacher/OT ratings of student proficiency.

Grade	IOWAN and Teacher Survey Ratings	IOWAN and OT Survey Ratings
Second	.54	.59
Fourth	.58	.55

Discussion: The concurrent validity coefficients generated by this study document a moderate to high level of concurrent validity between both a published, standardized measure of handwriting proficiency and a standardized rating scale. Most concurrent validity coefficients were in the moderate to high range.

Additional validity evidence is provided by earlier examinations of the IOWAN measure. Student performance on the IOWAN clearly discriminates special education students from general education students. It discriminates second graders' performance from fourth graders' performance and it appears to reflect developmental differences between girls and boys performances, especially as students progress into 4th grade.

