

The WIAT-II, Statistical Significance of Difference Scores, and Severe Discrepancy Ability-Achievement Models: A Discussion of Important Issues

prepared by
**George McCloskey, Ph.D., Senior Research Director,
The Psychological Corporation**

A number of clinicians have been asking questions about the Ability-Achievement Significant Difference Tables of the WIAT-II Scoring and Normative Supplement.

The most frequently voiced concern relates to the smaller point values required to denote a significant difference, especially for the predicted model. Most of the clinicians who express these concerns are familiar with the original WIAT Ability-Achievement Significant Difference tables and are accustomed to seeing point values of much larger magnitude denoting statistically significant differences.

Statistical Significance of the Difference Between Predicted Achievement and Actual Achievement

Some of the difference in tabled values is the result of improved reliability of some of the WIAT-II subtests (the better the reliability of the subtest, the smaller the difference required to reach statistical significance). However, the majority of the difference is attributable to a change in one aspect of the statistical formula used to calculate statistically significant differences between predicted and actual achievement scores.

The change in formula is discussed in the WIAT-II Updated Examiner's Manual on page 157, where it is stated: "Users of the WIAT will note the similarity of formulas presented here **but also that the calculations of extreme discrepancies (The Psychological Corporation, 1992, p. 188) are no longer included, in favor of the more widely used standard error of the residual.**

The Concepts of Statistical Significance and Severe Discrepancy

When using tables of statistical significance such as those provided in the WIAT-II Updated Scoring and Normative Supplement for Grades Pre-K-12 (Appendix I, Table I.2, pages 317-318), **it is important to keep in mind the distinction between statistical significance and severe discrepancy. The WIAT-II Ability-Achievement Discrepancy tables don't say anything about severe discrepancy.** They simply report the magnitude of difference required (at the .05 or .01 level) in order for a test user to say that there is a difference between the predicted and actual achievement scores that can't be explained simply by measurement error. The tables simply report the size of a difference between actual and predicted scores required to make the statement that, in fact, the difference is real in a statistical sense. Another way to think of it is that the confidence intervals of the predicted and actual scores don't overlap, so there is a real difference between these two scores.

Statistical significance as it is represented in Table I.2 is different from the concept of severe discrepancy. Establishing a severe discrepancy involves going beyond the basic calculation of statistically significant difference. There are various ways to operationally define the concept of severe discrepancy as it is used in Ability-Achievement discrepancy models. Many states vary considerably in how they do this. Indeed, many school systems within the same state vary widely in how they operationally define severe discrepancy.

It should be noted that using statistically significant differences of 9-15 points, such as those reported in the original WIAT manual, as acceptable criteria for an operational definition of severe discrepancy would identify an extremely large number of students as qualifying for special education. Such a modest criterion would easily identify 15% or more of the population as having severe discrepancies between predicted and actual achievement scores on any given subtest of the WIAT.

Frequency of Differences Between Predicted Achievement and Actual Achievement

One effective alternative to reliance on the statistically significant differences between predicted and actual scores would be to look not just at statistical significance, but also at the frequency of occurrence of score differences in the standardization and linking samples. For example, a score difference that occurred with less than X% of the standardization linking sample cases (see WIAT-II Updated Scoring and Normative Supplement for Grades Pre-K – 12, Appendix I, Tables I.7- I.11, pages 321-323) is one such criterion for a severe discrepancy.

Using a frequency of occurrence criterion of approximately 10% or less to establish a severe discrepancy, most subtest comparisons would require at least a 20-25 point difference between predicted and actual scores to say that a severe discrepancy is present. This seems a much more reasonable criterion for severe discrepancy than the difference required for statistical significance.

There is certainly good precedent for doing this. Note that the WIAT-II Updated Examiner's Manual on page 154 states:

"In addition, as recommended by assessment professionals such as Sattler (2001), the frequency of discrepancies in the norming or linking samples should be considered. In other words, "How often do discrepancies of this size occur?" ...An evaluation of statistical significance provides an answer to the concern that measurement error in each test may produce a score difference by chance. The difference must be of sufficient size to minimize the probability that a difference occurred because of unreliability or chance errors in the assessment. An evaluation of the frequency provides an answer to the concern that certain magnitudes of difference may not be rare in the normative population (estimated by the linking-sample frequency of the calculated discrepancies for each examinee). The differences should be sufficiently large and relatively rare in the population."

The Linking Sample percentages of differences obtained when comparing WISC-III and WIAT-II scores are reported in the WIAT-II Scoring and Normative Supplement for Grades Pre-K - 12, Appendix I, starting on page 321 in Tables I.7 through I.11. These tables could be used to effectively construct an operational definition of severe discrepancy that takes both statistical significance and frequency of occurrence of the size of a difference into consideration, thereby producing a more balanced perspective on the use of objective score data in educational decision-making processes.

As an example, consider the following:

CASE DATA:

Thomas CA: 8:8 Grade 3

WISC-III Verbal Comprehension Index Standard Score = 110

WIAT-II Word Reading Subtest Standard Score = 84

David CA: 8:8 Grade 3

WISC-III Verbal Comprehension Index Standard Score = 110

WIAT-II Word Reading Subtest Standard Score = 88

Mark CA: 8:8 Grade 3

WISC-III Verbal Comprehension Index Standard Score = 110

WIAT-II Word Reading Subtest Standard Score = 93

WIAT-II TABLE DATA:

1. WIAT-II Table H.7. WIAT-II Subtest and Composite Standard Scores Predicted from the WISC-III VCI Scores for Children Aged 6:0 – 16:11

Word Reading Subtest **Predicted Standard Score = 104**

2. WIAT-II Table I.2. Differences Between Predicted and Actual Subtest and Composite Standard Scores Required for Statistical Significance Using the WISC-III Scores: Ages 6:0 – 11:11

Word Reading Subtest **.05 Level = 7.13 .01 Level = 9.38**

3. WIAT-II Table I.10. Differences Between Predicted and Actual Subtest and Composite Standard Scores Obtained for Various Percentages of Children in the WISC-III Linking Sample Based on VCI Scores (selected percentage levels)

	20%	10%	5%	1%
Word Reading Subtest	10	15	19	27

CASE SCORE COMPARISONS:

	Actual	Predicted	Difference
Thomas	84	104	-20
David	88	104	-16
Mark	93	104	-11

DISCUSSION:

Note that Thomas, David and Mark all demonstrate Word Reading Subtest Standard Score differences between Actual and Predicted scores that are statistically significant at both the .05 and .01 levels.

When these differences are compared to the differences between predicted and actual scores obtained by the standardization sample, however, Thomas' score difference of 20 points exceeds the difference found in 5% or fewer of the standardization cases, while David's difference of 16 points exceeds the difference found in 10% or fewer of the standardization cases and Mark's score difference of 11 points exceeds only the difference found in 20% or fewer of the standardization cases.

Using the rationale presented above, the determination of whether or not any or all of these students meet the criterion of a severe discrepancy would depend on the degree of unusualness of occurrence of a difference in the standardization sample that was selected as an operational definition of severe discrepancy.

If the value were selected based on a criterion of equaling or exceeding the difference earned by 5% or fewer standardization cases (19 points), only Thomas would meet the criterion.

If the difference earned by 10% or fewer standardization cases (15 points) were selected, then both Thomas and David would meet the criterion for severe discrepancy.

In order for all three students to meet the criterion, then a difference earned by 20% or fewer of the standardization cases (10 points) would need to be selected as the criterion for severe discrepancy.

Conversely, if the 1% figure were selected (27 points), then none of these students would meet the criterion for severe discrepancy.