Technical Assistance Paper	309159

Standard Error of Measurement (SE*m*)

RATIONALE

The following technical assistance paper has been developed to address frequently raised questions about the use of the standard error of measurement in interpreting psychological test results, particularly tests of intelligence. This technical assistance paper has undergone substantial review by exceptional education personnel, school psychologists and psychologists, and nationally respected experts in psychological testing.

As with any professional decision made regarding data obtained under standardized testing conditions, the evaluator's professional interpretation of the results should be accepted as a reflection of the student's abilities at a given point in time. If the data interpretation cannot be accepted with a high level of confidence, the evaluator should secure additional data rather than to leave the data open to alternative interpretations by others who did not conduct the evaluation.

1. **What is the standard error of measurement?**

    The standard error of measurement (SE*m*) estimates how repeated measures of a person on the same instrument tend to be distributed around his or her "true" score. The true score is always an unknown because no measure can be constructed that provides a perfect reflection of the true score. SE*m* is directly related to the reliability of a test; that is, the larger the SE*m*, the lower the reliability of the test and the less precision there is in the measures taken and scores obtained. Since all measurement contains some error, it is highly unlikely that any test will yield the same scores for a given person each time they are retested.

2. **What is the confidence interval and how does the SE*m* relate to it?**
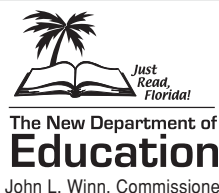
    Statements about an examinee's obtained score (the actual score that is received on a test) are couched in terms of a confidence interval — a band, interval, or range of scores that has a high probability of including the examinee's "true" score. Depending on the level of confidence one may want to have about where the "true" score may lie, the confidence band may be small or large. Most typical confidence intervals are 68%, 90%, or 95%. Respectively, these bands may be interpreted as the range within which a person's "true" score can be found 68%, 90%, or 95% of the time. It is not possible to construct a confidence interval within which an examinee's true score is absolutely certain to lie. It is important to report the confidence interval associated with a child's obtained score so that the reader can be informed of the probability that the examinee's true score lies within a given range of scores.

3. **What is the recommended confidence interval that should be used?**

    Selection of the confidence interval will depend on the level of certainty that one wishes to have about where the examinee's true score may lie given their obtained score. The 68% confidence level is the one most typically reported in psychoeducational evaluation reports. This is often reported in the following manner; "Given the student's obtained score of _____, there are two out of three chances that the individual's true score would fall between_____(low score in range) and_____(high score in range)." Most test authors, however, suggest that a higher level of confidence statement be made. These would be stated as "nine out of ten chances" for the 90

percent confidence band or "95 out of 100 chances" for the 95 percent confidence band. Obviously the increased levels of confidence would expand the range of scores included in the probability statements. Test publisher recommendations should be followed in reporting confidence levels.

4. **What other statistical information besides the SE*m* and confidence intervals is used in interpreting test scores? How are these choices made?**

One may choose, when such technical information is available, to use the SE*m* for a given age group, for a given ability group, or for all groups combined (the mean SE*m*). In addition, the standard error of estimation, when it is provided, may be the preferred measure of score precision. Given the fact that norm-referenced, standardized tests (not just tests of intelligence) provide technical and statistical properties in test manuals and reference materials, the selection of the type of statistical information that is used in interpreting obtained scores should be based on the recommendations of the test publisher.

5. **What is the standard error of estimation?**

Some tests, such as the *Wechsler Intelligence Scale for Children, 4th Edition* (WISC-IV), provide the user with the standard error of estimation (SE*est*) another form of standard error of measurement. This statistic takes into account regression toward the mean and the fact that scores at the extreme end (very high or very low scores) of the distribution are more prone to error than scores near the average. Because of this fact, the standard error of estimation is not equivalent around the obtained score. This would be reflected in a statement such as, "Given an obtained IQ score of 125, there are two out of three chances that this student's true score lies between 122 and 127 (-3 and +2 of the obtained score)." The WISC-IV manual provides standard error of estimation for the 90 and 95 percent levels of confidence.

6. **How should the SE*m* be used in program eligibility determination?**

The SE*m* is a characteristic of the test that reflects the probability that an examinee's true score falls within a given range of scores. No score within the range of scores (except the obtained score) has a higher probability of occurring than any other score within that range. Using the 68% confidence level, for example, if a child receives an intelligence test score of 115 with a SE*m* of three (3) points, there is a 68% probability that the child's true score falls within the range of 112 to 118. It would not be appropriate to select the highest or lowest numbers within that range as the best estimate of the child's true score. In fact, the best estimate of any child's true score on a given test is the obtained score given, appropriate test administration procedures are followed, there is good effort and motivation on the part of the examinee, and there are no conditions within the testing situation that would unduly influence test scores. In the sample case just cited, that would be 115. The SE*m* should be treated as information one has about a test to be considered by the examiner and/or eligibility committee in determining the presence or not of a disability or giftedness.

7. **What factors would indicate that the obtained score is not the best estimate of the child's abilities?**

Behaviors exhibited by the child during evaluation may call into question the confidence one may have in the obtained score. These behaviors should be well cited by the examiner. Examples might include a child's need for frequent rest breaks during testing due to inattention or fatigue, or minor, but noticeable difficulties a child may have in manipulating test materials. These behaviors would not be severe enough to invalidate the test results, but may have a modest influence on the performance of the child. Background information about the child may also be considered in decisions about the confidence that can be placed in the obtained score. If the child had been tested at an earlier date on the same instrument and revealed a modest recollection of the material, the obtained score may reflect the previous exposure to the test materials.

8. **Are there specific actions that should be taken by the evaluator in deciding how to obtain a measure of or to estimate the child's ability if the obtained score is judged to not be the best estimate?**

The examiner should have sufficient basis for dismissing the obtained score as the best estimate of a child's true score on a given test. The standard error of measurement should not be used to unilaterally extend or restrict the definition of giftedness or a disability such as mental retardation. That is, scores that fall within a given confidence interval of two standard deviations above the mean or two standard deviations below the mean are not "automatically" interpreted as "gifted" or "mentally handicapped" respectively. The eligibility committee should guard against promoting such actions as, there is just as much of a chance of inaccurately placing a student, as there is in inaccurately not placing a student in a program. As stated earlier, all things being equal, the best estimate of the true score for a given individual is always the obtained score. Naturally, additional testing is indicated if very questionable results are achieved.

9. **Is it appropriate to use the SE*m* when determining program eligibility?**

Yes, but only under limited conditions. SBER 6A-6.0331(l)(b), FAC, states in part, "The district's evaluation procedures shall provide for the use of valid tests and evaluation materials, administered and interpreted by trained personnel, in conformance with instructions provided by the producer of the tests or evaluation materials." SBER 6A-6.03019(2)(a)3, FAC, states that eligibility for the gifted program requires "(S)uperior intellectual development as measured by an intelligence quotient of two (2) standard deviations or more above the mean on an individually administered standardized test of intelligence." Intellectual development can be measured through the administration of a standardized test of intelligence (such as the WISC-IV or Stanford Binet-V) where an obtained score is generated. This obtained score is used to determine program eligibility. If the obtained score falls at or above two standard deviations above the mean, the student would meet this aspect of program eligibility. Eligibility committees may also judge a student eligible if there is overwhelming evidence in favor of more liberal interpretation, that is, by considering the standard error of measurement. However, one should not make a standard practice of adding the SE*m* to the child's obtained score and deeming the child eligible for the gifted program if the added points result in a number that is at or above two standard deviations above the mean. Conversely, one would not routinely subtract the SE*m* from a student's score at the lower end of the distribution to deem the student eligible for the mentally handicapped program. Whenever questions arise about specific test scores, the individual who conducted the evaluation should be consulted

10. **How should test scores be provided in interpretative reports?**

In order to accurately reflect the measurement error inherent in each test, it is appropriate to reflect in test interpretation reports the confidence that one has regarding the range within which an examinee's true score lies given their obtained score. Although the obtained score is the best estimate of the true score, one can be more confident in stating that the true score lies within a given range of the obtained score. Recommended confidence intervals are generally provided in test manuals.

11. **Must the standard error of measurement be reported for all test scores?**

No. This is but one of many statistical features of a test and does not have to be reported by the examiner in an interpretive report. Because of the greater weight that is placed on scores obtained on tests of intelligence, it is strongly recommended that appropriate confidence levels be provided to the reader of the psychological report. For other types of tests, the examiner should be prepared to provide this information, particularly when it is critical to decisions that must be made about the examinee.