

Applying and Interpreting the Standard Error of Measurement and Standard Error of the Estimate in Intelligence Testing

Gordon E Taub^{*1} Oliver E. Edwards²

¹Department of Counselor Education and School Psychology University of Central Florida, USA

²Department of Counselor Education and School Psychology, University of Central Florida, USA

***Corresponding author:** Gordon E Taub, Department of Counselor Education and School Psychology University of Central Florida, USA

Citation: Taub GE. (2020) Applying and Interpreting the Standard Error of Measurement and Standard Error of the Estimate in Intelligence Testing. *Forensic Stud*: 03: 117. DOI: 10.29011/2577-1523.100017

Received Date: 18 November, 2019; **Accepted Date:** 25 February, 2020; **Published Date:** 28 February, 2020

Abstract

Many education, legal, and psychological professionals are unaware of the difference between the standard error of measurement (SEM) and the standard error of the estimate (SEE). Confusion between the SEM and the SEE may be due, in part, to the way they are presented and applied in intelligence test manuals. Some test publishers provide users with an instrument's SEM for the standardization sample and then include tables in their tests containing confidence intervals derived from the SEE. It is important for psychological professionals who report and interpret test results to know the difference between the SEM and the SEE and when each should be used.

Introduction

An examinee's score on a test of intelligence is relatively easy to understand, it is simply the individual's observed Full-Scale Intelligence Quotient (FSIQ). However, intelligence tests are not perfect or infallible instruments, there is error associated with a person's observed FSIQ score. This measurement error is reflected in a test's reliability. The relationship between a test's error and reliability is inverse. A test with high reliability will have less error. In contrast, a test with low reliability will have more error. An FSIQ score from a test with high reliability will have less error around an observed FSIQ score.

One theoretical way to see how much error is associated with an observed FSIQ is to have an individual take an intelligence test an infinite number of times and average the scores across all administrations. The average of these scores is the best estimate of the person's *true FSIQ* score. The scores above and below the true score represent the error associated with the instrument.

Standard Error of Measurement

Obviously, it is not possible to obtain a person's true FSIQ score this way because we are not able to administer an intelligence test an infinite number of times to the same individual holding testing effects, fatigue, maturation, etc. constant. Another method is to assume the observed FSIQ score is the best estimate of the person's intelligence, while also knowing there is error associated

with the observed FSIQ score. Using the instrument's reliability coefficient, it is possible to account for the error inherent within the instrument. This error is accounted for via the standard error of measurement (SEM). An instrument with relatively high reliability will have a relatively small SEM, whereas an instrument with lower reliability will have a larger SEM because there is more error.

The SEM is used to create a range of scores around an observed FSIQ score. Reporting FSIQ scores as a range of scores provides a degree of confidence that the person's true FSIQ score is contained within the reported range of scores [1]. The SEM is derived from a test's standard deviation and can be expressed in SEM units. For example, an SEM of 3 means an examinee's true score lies within 3 points above and 3 points below the observed FSIQ score. Because the SEM is based on a normal curve, we can be 68% confident that the examinee's true FSIQ score lies within the range of scores between 3 FSIQ points above and below the observed FSIQ score.

To be 95% confident the true FSIQ score lies within a reported range of scores requires using 1.96 SEM. For example, with an SEM of 3 this is obtained by multiplying $3 \times 1.96 = 5.88$ FSIQ points or rounded to 6 FSIQ points. This means, we have 95% confidence the individual's true FSIQ score is somewhere between 79-91 (85 - 6 and 85 + 6). Psychological professionals provide confidence intervals via the SEM to indicate the individual's FSIQ is actually a range of scores (not a single score) and within this

range is the examinee's true FSIQ score [2]. For 68% confidence the true FSIQ is represented by a range of scores it is necessary to report confidence intervals of ± 1 SEM of the observed FSIQ score. To have 95% confidence, the FSIQ is reported as ± 1.96 SEM, and for 99% confidence FSIQ scores are reported as a range of scores ± 2.58 SEM. It is interesting to note that a test which is 100% reliable will have an SEM of 0 and for a test with 0 reliability, the SEM will be the test's standard deviation.

Standard Error of the Estimate

Many education, legal, and psychological professionals assume the SEM is the same as the standard error of the estimate (SEE). The SEM and the SEE are not the same. The SEE accounts for regression to the mean. Regression to the mean is a statistical phenomenon wherein random variables that are far from the mean (higher or lower), will over time, tend to move closer to the mean. When calculating the SEE for an observed FSIQ score, the formula *transforms* the observed FSIQ score into a score that is closer to the mean or average of the instrument. The formula then calculates the error around this *new* FSIQ score. It is important to note, the observed FSIQ score is used when applying the SEM and the regressed or new FSIQ score is used when calculating the SEE. The SEE is calculated using regression analysis, this means the SEE is used to predict something that has not occurred. For example, in intelligence testing the SEE may be used to predict a future score on an intelligence test.

When using the SEE, FSIQ scores that are far from the mean will be regressed (changed) more than FSIQ scores that are closer to the test's average score. When using the SEE with FSIQ scores that are much higher and lower than the mean, the range of scores will be asymmetrical. For example, the SEE for an FSIQ score in the gifted range (i.e., FSIQ 130) on the Wechsler Adult Intelligence Scale-Fourth Edition [3] will be 125-133 or 5 FSIQ

points below and 3 points above the examinee's FSIQ score. The direction of the skew will always favor the mean of the instrument, due to regression to the mean.

The publisher of the Wechsler Intelligence scales provides the SEM for the standardization sample and the SEM for each age-level. However, the tables in the test's manual present confidence intervals calculated using the SEE. Thus, the SEE in the Wechsler manuals are centered on the new regressed FSIQ score. Nevertheless, the publisher notes, "practitioners may wish to calculate confidence intervals centered on the observed score" [4]. Meaning, practitioners may find applying the SEM to the observed FSIQ score is more appropriate than using the publisher's tables developed using SEE.

The decision to report either the SEM or the SEE depends on the purpose of the evaluation. If a practitioner is interested in identifying and accounting for an individual's true FSIQ score, such as when determining program eligibility, reporting confidence intervals around the observed FSIQ score using SEM is most appropriate. If on the other hand, if the practitioner is interested in predicting a score on a future test administration, reporting confidence intervals using the SEE may be more appropriate.

References

1. American Educational Research Association (2014). *Standards for educational and psychological testing*. Washington, DC, USA.
2. Wechsler D (2008a) Wechsler Adult Intelligence Scale-Fourth Edition: Technical and Interpretive Manual. Pearson Assessment. USA.
3. Wechsler D (2008b) Wechsler Adult Intelligence Scale-Fourth Edition. Pearson Assessment, USA.
4. Wechsler, D. (2014). Wechsler Intelligence Scale for Children—Fifth Edition: Technical and Interpretive Manual. NCS Pearson. USA.