Educational Measurement
ISSUES AND PRACTICE

NCME national council on measurement in education

Educational Measurement: Issues and Practice
Fall 2014, Vol. 33, No. 3, pp. 55–56

A Simple Equation to Predict a Subscore's Value

Richard A. Feinberg and Howard Wainer, *National Board of Medical Examiners*

Subscores are often used to indicate test-takers' relative strengths and weaknesses and so help focus remediation. But a subscore is not worth reporting if it is too unreliable to believe or if it contains no information that is not already contained in the total score. It is possible, through the use of a simple linear equation provided in this note, to determine if a particular subscore adds enough value to be worth reporting.

Keywords: subscores, reliability, value added, orthogonally, proportional reduction in mean squared error

For at least the last 4,000 years people have taken tests. Historically, tests were contests with the winners getting the job, being admitted to the school, or getting awarded the prize. Secondarily tests were also used as prods: "Why is Linda studying?" "She has a test." And their most modern use is as a measuring instrument. It is this latter use with which this note is concerned.

Most would agree that a test that may take an examinee hours to complete is being inefficiently used if all that it yields is a single number. Surely, if there is more information to be mined from all of that effort than what can be represented in a single number, it should be reported. If not, do we really need all that time and effort to obtain a single summary number that is accurate enough for the purposes at hand?

Thus, it is natural to want to augment the single overall score with a number of subscores that give more specific insights into the examinee's knowledge and abilities. And so, as surely as day follows night, the subscores generated must satisfy two minimal conditions to justify their existence:

1. the subscore must be reliable enough so that those who use it are not chasing noise. The more reliable the subscore the greater its potential usefulness.
2. The subscore must contain information that is not available from the rest of the test. The more orthogonal the subscore is to the rest of the test, the greater its potential value.

One can think of a subscore's value as being the accuracy with which it can predict a parallel measure of itself in some future assessment. If it can do so better than the total test score, the subscore adds value; if it cannot, it does not. Shelby Haberman (2008) calculated statistics for evaluating precision of the subscore and total score as predictors, that when compared provide a measure of the marginal value that a subscore adds in this task.

From Haberman's statistic it is easy to calculate the proportional reduction in mean squared error (PRMSE). If the ratio of these statistics, subscore over the total score (what we refer to as a value added ratio), is less than one, it is less valuable than total score; greater than one, of greater value.

Haberman's formulas make a valuable contribution, because once a test is built and subscores are constructed we can calculate which of those subscores deserve being reported and which ones do not.

Of course, it would be even more valuable if we could know this information in advance. Fortunately, if we can calculate just two elementary summary statistics for every proposed subscore, we can predict with high accuracy, the added value of a proposed subscore. To do this we need two components:

1. The reliability of the subscore—let us call this r_1.
2. The extent of the subscore's orthogonality relative to the total test score; that is the disattenuated correlation of the subscore with the total test score without the items in the subscore. Specifically, the raw correlation between the subscore and remainder of the test divided by the square root of the product of their reliabilities—let us call this r_2.

With these two statistics in hand the estimate of that subscore's value added can be calculated from

$$\text{Value Added Ratio} = 1.15 + 0.51r_1 - 0.67r_2. \tag{1}$$

Equation 1 was derived from the results of an extensive simulation (Feinberg, 2012) and was validated, at least in part, by noting the remarkably close match to the empirical results reported by Sinharay (2010).

To answer the question whether subscores have any value beyond that communicated by the total score, one need only:

i. Divide up a test into its projected subscores;
ii. Calculate the reliability of each subscore;
iii. Calculate the disattenuated correlation of each subscore with its remainder score; and
iv. Insert the results of steps (ii) and (iii) into Equation 1 and see which, if any, of the subscores yield a value added greater than one.

If there are none (as Sinharay (2010) found), then there is no evidence to support the reporting of subscores.

This work is collaborative and the order of authorship is alphabetical.

Table 1. Worked Example

	Statistic	Subscore 1	2	3	4	5	6
Test 1	Subscore Reliability	.86	.59	.73	.73	.75	.87
	Remainder Score Reliability	.93	.93	.92	.93	.92	.92
	Raw Correlation	.37	.61	.80	.69	.77	.62
	Disattenuated Correlation	.41	.82	.97	.84	.92	.69
	PRMSE	.44	.68	.89	.74	.85	.66
	Subscore Reliability/PRMSE	2.0	.9	.8	1.0	.9	1.3
	Equation (1)	1.3	.9	.9	1.0	.9	1.1
	Same Decision?	Y	Y	Y	Y	Y	Y
Test 2	Subscore Reliability	.78	.63	.78	.58	.56	.54
	Remainder Score Reliability	.88	.90	.88	.90	.90	.90
	Raw Correlation	.69	.62	.70	.61	.62	.60
	Disattenuated Correlation	.83	.82	.85	.84	.87	.86
	PRMSE	.74	.70	.76	.71	.61	.68
	Subscore Reliability/PRMSE	1.1	.9	1.0	.8	.9	.8
	Equation (1)	1.0	.9	1.0	.9	.9	.8
	Same Decision?	Y	Y	Y	Y	Y	Y

Technical Appendix

Equation 1 can be validated from Table 2 in Sinharay (2010); specifically, by using the two subscore condition where the remainder of the test is the secondary subscore. A worked example is shown below to illustrate the methodology. Table 1 includes all the necessary information for computing the result specified in Equation 1 and, for comparison, Haberman's PRMSE. Test 1 is a residency in-training exam and Test 2 is a public health credentialing exam; both have 6 subscores. Note that, for the purpose of Equation 1, disattenutated correlation is calculated as

$$r_2 = \frac{\text{Raw Correlation}}{\sqrt{\text{Subscore Reliability} \times \text{Remainder Reliability}}}.$$

Subscore value for the second subscore on Test 1, calculated through Haberman's methodology:

$$\text{Value Added Ratio} = \frac{\text{Subscore Reliability}}{\text{PRMSE}} = \frac{0.59}{0.68} = 0.9.$$

Using Equation 1:

$$\text{Value Added Ratio} = 1.15 + 0.51r_1 - 0.67r_2 = 1.15 + (0.51)(0.59) - (0.67)(0.82) = 0.9.$$

In all cases, both approaches yield the same decisions, although there are minor differences in the outcome statistic due to the approximate nature of the regressed estimate.

References

Feinberg, R. A. (2012). *A simulation study of the situations in which reporting subscores can add value to licensure examinations* (PhD dissertation). Retrieved October 31, 2012, from ProQuest Digital Dissertations database (UMI No. 3526412).

Haberman, S. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics*, *33*, 204–229.

Sinharay, S. (2010). How often do subscores have added value? Results from operational and simulated data. *Journal of Educational Measurement*, *47*, 150–174.