Research Report

Subscores and Validity

Shelby J. Haberman

December 2008

ETS RR-08-64

Listening. Learning. Leading.®

Subscores and Validity

Shelby J. Haberman
ETS, Princeton, NJ

December 2008

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS's constituents and the field.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Abstract

In educational testing, subscores may be provided based on a portion of the items from a larger test. One consideration in evaluation of such subscores is their ability to predict a criterion score. Two limitations on prediction exist. The first, which is well known, is that the coefficient of determination for linear prediction of the criterion score by the subscore cannot exceed the reliability coefficient of the subscore. The second limitation is on incremental validity. The coefficient of determination for linear prediction of the criterion score by both the total score and the subscore is at least as great as the coefficient of determination for linear prediction of the criterion score by only the total score. Incremental validity may be measured by the difference between these two coefficients of determination. This difference is no greater than the reliability of the residual from linear prediction of the subscore by the total score.

Key words: Partial correlation, reliability

Acknowledgments

This article has benefitted from conversations with Sandip Sinharay and from comments by Neil Dorans and Michael Walker.

When subscores based on sections of a larger test are employed to provide more detailed information about an examinee than is available from a total test score, it is reasonable to ask to what extent the subscores can provide useful predictions of a criterion score. A more subtle question is whether the subscore can provide incremental validity given that the total score is already used to predict the criterion score. In fact, significant limitations exist on validity. These limitations depend on the reliability of the subscore, the reliability of the total score, and on the correlation of the true subscore and true total score. The limitations apply to any criterion score.

To examine these limitations, some basic results from classical test theory are provided in section 1. In section 2, these result are provided to yield the desired limits. Some examples with operational data are provided in section 3. In section 4, conclusions are reached concerning practical implications. The notation and arguments used parallel those in Haberman (2008).

1 Results From Classical Test Theory

Let S_X be a random variable that represents an observed subscore of an examinee randomly selected from some population, and let S_Z be a random variable that represents the observed total score for that randomly selected examinee. Let $S_Y = S_Z - S_X$, the observed remainder score, be the portion of the observed total score not ascribed to the observed subscore. Let S_V be a random variable that represents the observed value of an external criterion score for the examinee. This report considers linear prediction of S_V by S_X, linear prediction of S_V by S_Z, and linear prediction of S_V by both S_X and S_Z. As in classical test theory (Haberman, 2008; Holland & Hoskens, 2003), for the randomly selected examinee, consider a randomly selected test form from a collection of parallel test forms and a randomly selected validity measurement from a corresponding collection of parallel measurements of the validity criterion under consideration. Assume that selection of test form and criterion score are independent. Then $S_X = \tau_X + e_X$, $S_Y = \tau_Y + e_Y$, $S_Z = \tau_Z + e_Z$, and $S_V = \tau_V + e_V$. Here the true subscore τ_X is the conditional expected value of S_X given the examinee, $e_X = S_X - \tau_X$ is the measurement error of the subscore, the true total score τ_z is the conditional expected value of S_Z given the examinee, $e_Z = S_Z - \tau_Z$ is the measurement error of the total score, the true remainder score τ_Y is the conditional expected value of S_Y given the examinee, $e_Y = S_Y - \tau_Y$ is the measurement error of the remainder score, the true criterion score τ_V is the conditional expected value of S_V given the examinee, and $e_V = S_V - \tau_V$ is the measurement error of the criterion score. Under the assumption that the observed measurements

S_X, S_Z, and S_V all have finite and positive variances, the observed subscore S_X, the true subscore τ_X, the measurement error e_X, the true total score τ_Z, the error of measurement e_Z, the observed remainder score S_Y, the true remainder score τ_Y, the measurement error e_Y, the true criterion score τ_V, and the corresponding measurement error e_V are all random variables with finite means and variances. The expectations satisfy the constraints that $E(S_X) = E(\tau_X)$, $E(S_Z) = E(\tau_Z)$, $E(S_Y) = E(\tau_Y)$, $E(S_V) = E(\tau_V)$, and $E(e_X) = E(e_Z) = E(e_Y) = E(e_V) = 0$. The errors of measurement are all uncorrelated with the true scores. The error of measurement e_X for the subscore and the error of measurement e_Y for the remainder score are uncorrelated, so that the covariance $\mathrm{Cov}(e_X, e_Z) = \mathrm{Cov}(e_X, e_X + e_Y)$ of the measurement errors e_X and e_Y is the variance $\sigma^2(e_X)$ of the error of measurement e_X of the subscore (Haberman).

To avoid trivial cases, it is assumed that the variance $\sigma^2(\tau_X)$ of the true subscore τ_X, the variance $\sigma^2(\tau_Y)$ of the true remainder score, the variance $\sigma^2(\tau_Z)$ of the true total score τ_Z, the variance $\sigma^2(\tau_V)$ of the true criterion score, the variance $\sigma^2(e_X)$ of the measurement error e_X, the variance $\sigma^2(e_Y)$ of the remainder score e_T, and the variance $\sigma^2(e_V)$ of the measurement error e_V are all positive. The variance $\sigma^2(e_Z) = \sigma^2(e_X) + \sigma^2(e_Y)$ of the measurement error e_Z, the variance $\sigma^2(S_X) = \sigma^2(\tau_X) + \sigma^2(e_X)$ of the observed subscore S_X, the variance $\sigma^2(S_Y) = \sigma^2(\tau_Y) + \sigma^2(e_Y)$ of the observed remainder score, the variance $\sigma^2(S_Z) = \sigma^2(\tau_Z) + \sigma^2(e_Z)$ of the observed total score S_Z, and the variance $\sigma^2(S_V) = \sigma^2(\tau_V) + \sigma^2(e_V)$ of the observed criterion score S_V are all positive (Lord & Novick, 1968, p. 57). Thus the observed subscore S_X has reliability coefficient $\rho^2(S_X, \tau_X) = \sigma^2(\tau_X)/\sigma^2(S_X)$ equal to the square of the correlation $\rho(S_X, \tau_X)$ of S_X and the true subscore τ_X, the observed remainder score S_Y has reliability coefficient $\rho^2(S_Y, \tau_Y) = \sigma^2(\tau_Y)/\sigma(S_Y)$ equal to the square of the correlation $\rho(S_Y, \tau_Y)$ of S_Y and the true remainder score τ_Y, the observed total score S_Z has reliability coefficient $\rho^2(S_Z, \tau_Z) = \sigma^2(\tau_Z)/\sigma^2(S_Z)$ equal to the square of the correlation $\rho(S_Z, \tau_Z)$ of S_Z, and the true total score τ_Z, and the observed criterion score S_Z has reliability coefficient $\rho^2(S_V, \tau_V) = \sigma^2(\tau_V)/\sigma^2(S_V)$ equal to the square of the correlation $\rho(S_V, \tau_V)$ of S_V and the true criterion score τ_V. All reliability coefficients are positive and less than 1 (Lord & Novick, p. 61).

2 Prediction of the Criterion Score

If $\rho(S_X, S_V)$ is the correlation of the observed subscore S_X and the observed criterion score S_V and if $\rho(\tau_X, \tau_V)$ is the correlation of the true subscore τ_X and the true criterion score τ_V, then

$$\rho(S_X, S_V) = \rho(\tau_X, \tau_V)\rho(S_V, \tau_V)\rho(S_X, \tau_X) \tag{1}$$

(Holland & Hoskens, 2003), so that

$$|\rho(S_X, S_V)| \leq \rho(S_V, \tau_V)\rho(S_X, \tau_X) \tag{2}$$

and the coefficient of determination $\rho^2(S_V|S_X) = \rho^2(S_X, S_V)$ for prediction of S_V by S_X satisfies

$$\rho^2(S_V|S_X) \leq \rho^2(S_V, \tau_V)\rho^2(S_X, \tau_X).$$

The product $\sigma^2(S_V)\rho^2(S_V|S_X)$ is the mean-squared error achieved by linear prediction of the criterion score S_V by the observed subscore S_X. As is well known, (1) and (2) show that the ability to predict the validity criterion is constrained by the reliability coefficients of both the validity criterion and the subscore (Lord & Novick, 1968, p. 72). If the subscore has limited reliability, then prediction of the validity criterion cannot be very effective. Thus a subscore with a reliability of 0.25 cannot produce a coefficient of determination greater than 0.25, and the combination of a subscore with a reliability of 0.25 and a criterion score with a reliability of 0.25 cannot yield a coefficient of determination greater than 0.0625.

The incremental contribution of the observed subscore appears to have been less studied. To determine this contribution, one first considers linear prediction of the observed subscore S_X by the observed total score S_Z. The best linear predictor

$$L_{X \cdot Z} = E(S_X) + \beta_{X \cdot Z}[S_Z - E(S_Z)],$$

where

$$\beta_{Y\dot{X}} = \frac{\mathrm{Cov}(S_X, S_Z)}{\sigma^2(S_Z)}$$

is the regression coefficient for linear prediction of S_X by S_Z. The prediction error is then the residual subscore

$$S_{X \cdot Z} = S_X - L_{X \cdot Z}.$$

Because the observed residual subscore $S_{X \cdot Z}$ is uncorrelated with the observed total score S_Z, the coefficient of determination $\rho^2(S_V|S_Z, S_X)$ for prediction of the observed criterion score S_V by both the observed subscore S_X and the observed total score S_Z satisfies

$$\rho^2(S_V|S_Z, S_X) = \rho^2(S_V|S_Z) + \rho^2(S_V|S_{X\cdot Z})$$

(Lord & Novick, 1968, p. 266). A measure of the incremental validity of the subscore S_X is the difference

$$\rho^2(S_V|S_Z, S_Z) - \rho^2(S_V|S_Z) = \rho^2(S_V|S_{X\cdot Z})$$

between the coefficient of determination for prediction of the observed criterion score S_V by both the observed subscore S_X and the observed total score S_Z and the coefficient of determination for prediction of the observed criterion score S_V by the observed total score S_Z.

A bound on the coefficient of determination $\rho^2(S_V|S_{X\cdot Z})$ can be obtained by noting that the residual score $S_{X\cdot Z}$, which is a linear combination of the observed subscore S_X and the observed total score S_Z, has a true residual score

$$\tau_{X\cdot Z} = \tau_X - E(S_X) - \beta_{X\cdot Z}[\tau_Z - E(S_Z)],$$

a measurement error

$$e_{X\cdot Z} = e_X - \beta_{X\cdot Z} e_Z,$$

and a coefficient of reliability

$$\rho^2(S_{X\cdot Z}, \tau_{X\cdot Z}) = \frac{\sigma^2(\tau_{X\cdot Z})}{\sigma^2(S_{X\cdot Z})} = 1 - \frac{\sigma^2(e_{X\cdot Z})}{\sigma^2(S_{X\cdot Z})}.$$

Thus

$$\rho^2(S_V|S_{X\cdot Z}) \leq \rho^2(S_V, \tau_V)\rho^2(\tau_V, \tau_{X\cdot Z})\rho^2(S_{X\cdot Z}, \tau_{X\cdot Z}) \leq \rho^2(S_V, \tau_V)\rho^2(S_{X\cdot Z}, \tau_{X\cdot Z}).$$

The reliability $\rho^2(S_{X\cdot Z}, \tau_{X\cdot Z})$ is readily determined from sample data (Haberman, 2008). The variance

$$\sigma^2(S_{X\cdot Z}) = \sigma^2(S_X)[1 - \rho^2(S_X, S_Z)]$$

and the regression coefficient $\beta_{X\cdot Z}$ are estimated as in standard regression analysis. Classical reliability estimation methods lead to estimates of the variances of measurement $\sigma^2(e_X)$ and $\sigma^2(e_Y) = \sigma^2(e_Z) - \sigma^2(e_X)$. The decomposition $e_Z = e_X + e_Y$ leads to the formula

$$e_{X\cdot Z} = (1 - \beta_{X\cdot Z})e_X - \beta_{X\cdot Z} e_Y.$$

Because the measurement errors e_X and e_Y are uncorrelated, the variance $\sigma^2(e_{X\cdot Z})$ of the error of measurement $e_{X\cdot Z}$ satisfies

$$\sigma^2(e_{X\cdot Z}) = (1 - \beta_{X\cdot Z})^2\sigma^2(e_X) + \beta_{X\cdot Z}^2\sigma^2(e_Y).$$

Thus $\sigma^2(e_{X\cdot Z})$ is estimated by use of the estimates for $\beta_{X\cdot Z}$, $\sigma^2(e_X)$, and $\sigma^2(e_Y)$. In turn, the estimate for $\sigma^2(e_{X\cdot Z})$ and the estimate for $\sigma^2(S_{X\cdot Z})$ lead to an estimate for the reliability $\rho^2(S_{X\cdot Z}, \tau_{X\cdot Z})$.

As is evident from the examples in section 3, in many typical cases, the reliability of the residual subscore $S_{X\cdot Z}$ is low. The basic issue arises in the typical case in which the the true subscore S_X and the true remainder score τ_Y have a positive correlation $\rho(\tau_X, \tau_Y)$. Because the observed total score S_Z is the sum $S_X + S_Y$ of the observed subscore S_X and the observed remainder score S_Y,

$$\mathrm{Cov}(S_X, S_Z) = \mathrm{Cov}(S_X, S_Y) + \sigma^2(S_X),$$

$$\mathrm{Cov}(S_Y, S_Z) = \mathrm{Cov}(S_X, S_Y) + \sigma^2(S_Y),$$

and

$$\sigma^2(S_Z) = \sigma^2(S_X) + \sigma^2(S_Y) + 2\,\mathrm{Cov}(S_X, S_Y).$$

Thus

$$\beta_{X\cdot Z} = \frac{\mathrm{Cov}(S_X, S_Y) + \sigma^2(S_X)}{\sigma^2(S_X) + \sigma^2(S_Y) + 2\,\mathrm{Cov}(S_X, S_Y)}.$$

In like manner, if $\beta_{Y\cdot Z}$ is the regression coefficient for linear prediction of the observed remainder score S_Y by the total score S_Z, then

$$\beta_{Y\cdot Z} = \frac{\mathrm{Cov}(S_Y, S_Z)}{\sigma^2(S_Z)} = \frac{\mathrm{Cov}(S_X, S_Y) + \sigma^2(S_Y)}{\sigma^2(S_X) + \sigma^2(S_Y) + 2\,\mathrm{Cov}(S_X, S_Y)}.$$

As expected from the decomposition $S_X + S_Y = S_Z$, one has

$$\beta_{X\cdot Z} + \beta_{Y\cdot Z} = 1.$$

Because the true scores τ_X and τ_Y are uncorrelated with the measurement errors e_X and e_Y and because the measurement errors e_X and e_Y are uncorrelated,

$$\mathrm{Cov}(S_X, S_Y) = \mathrm{Cov}(\tau_X, \tau_Y)$$

(Lord & Novick, 1968, p.62). It follows that the observed subscore S_X and the observed remainder score S_Y are positively correlated if the true subscore τ_X and true remainder score τ_Y are positively correlated. In this case of positive correlation, the regression coefficients $\beta_{X \cdot Z}$ and $\beta_{Y \cdot Z}$ are both positive. This result has substantial impact when one considers the behavior of the variance components $\sigma^2(\tau_{X \cdot Z})$ and $\sigma^2(e_{X \cdot Z})$ that determine the reliability of the residual subscore.

The decompositions $\tau_Z = \tau_X + \tau_Y$ and $\beta_{X| \cdot Z} + \beta_{Y \cdot Z} = 1$ imply that the true residual subscore

$$\tau_{X \cdot Z} = \beta_{Y \cdot Z}[\tau_X - E(S_X)] - \beta_{X \cdot Z}[\tau_Y - E(S_Y)],$$

so that the variance

$$\sigma^2(\tau_{X \cdot Z}) = \beta^2_{Y \cdot Z}\sigma^2(\tau_X) - 2\beta_{Y \cdot Z}\beta_{X \cdot Z}\sigma(\tau_X)\sigma(\tau_Y)\rho(\tau_X, \tau_Y) + \beta^2_{X \cdot Z}\sigma^2(\tau_Y).$$

The correlation assumption implies the inequality

$$\sigma^2(\tau_{X \cdot Z}) < \beta^2_{Y \cdot Z}\sigma^2(\tau_X) + \beta^2_{X \cdot Z}\sigma^2(\tau_Y).$$

The previous formula for $\sigma^2(e_{X \cdot Z})$ can also be written as

$$\sigma^2(e_{X \cdot Z}) = \beta^2_{Y \cdot Z}\sigma^2(e_X) + \beta^2_{X \cdot Z}\sigma^2(e_Y),$$

and the variance

$$\sigma^2(S_{X \cdot Z}) = \beta^2_{Y \cdot Z}\sigma^2(\tau_{X \cdot Z}) + \beta^2_{X \cdot Z}\sigma^2(e_{X \cdot Z}).$$

The reliability formulas $\sigma^2(\tau_X) = \sigma^2(S_X)\rho^2(S_X, \tau_X)$ and $\sigma^2(\tau_Y) = \sigma^2(S_Y)\rho^2(S_Y, \tau_Y)$ and the decompositions $\sigma^2(S_X) = \sigma^2(\tau_X) + \sigma^2(e_X)$ and $\sigma^2(S_Y) = \sigma^2(\tau_Y) + \sigma^2(e_Y)$ imply that

$$\begin{aligned}
\sigma^2(\tau_{X \cdot Z}) &= \beta^2_{Y \cdot Z}\sigma^2(S_X)\rho^2(S_X, \tau_X) \\
&\quad + \beta^2_{X \cdot Z}\sigma^2(S_Y)\rho^2(S_Y, \tau_Y) \\
&\quad - \beta_{Y \cdot Z}\beta_{X \cdot Z}\sigma(S_X)\sigma(S_Y)\rho(S_X, \tau_X)\rho(S_Y, \tau_Y)\rho(\tau_X, \tau_Y)
\end{aligned}$$

and

$$\begin{aligned}
\sigma^2(S_{X \cdot Z}) &= \beta^2_{Y \cdot Z}\sigma^2(S_X) + \beta^2_{X \cdot Z}\sigma^2(S_Y) \\
&\quad - \beta_{Y \cdot Z}\beta_{X \cdot Z}\sigma(S_X)\sigma(S_Y)\rho(S_X, \tau_X)\rho(S_Y, \tau_Y)\rho(\tau_X, \tau_Y).
\end{aligned}$$

Thus the reliability coefficient $\rho^2(S_{X \cdot Z}, \tau_{X \cdot Z}) = \sigma^2(\tau_{X \cdot Z})/\sigma^2(S_{X \cdot Z})$ is less than the weighted average

$$\frac{\beta^2_{Y \cdot Z}\sigma^2(S_X)\rho^2(S_X, \tau_X) + \beta^2_{X \cdot Z}\sigma^2(S_Y)\rho^2(S_Y, \tau_Y)}{\beta^2_{Y \cdot Z}\sigma^2(S_X) + \beta^2_{X \cdot Z}\sigma^2(S_Y)}.$$

Obviously the reliability of the residual subscore $S_{X \cdot Z}$ will be low unless the subscore S_X or remainder score S_Y has high reliability; however, even if S_X and S_Y have high reliability, the reliability of $S_{X \cdot Z}$ is low if the correlation $\tau(\tau_X, \tau_Y)$ is high, for the negative term $-\beta_{Y \cdot Z}\beta_{X \cdot Z}\sigma(S_X)\sigma(S_Y)\rho(S_X, \tau_X)\sigma(\tau_Y)\rho(\tau_X, \tau_Y)$ then results in a substantial reduction in the reliability of $S_{X \cdot Z}$.

Some consideration of limits may help clarify results. For fixed $\rho(\tau_X, \tau_Y) < 1$, $\sigma^2(S_X)$, and $\sigma^2(S_Y)$, let the reliability coefficients $\rho^2(S_X, \tau_X)$ and $\rho^2(S_Y, \tau_Y)$ both approach 1. Then the reliability coefficients of both the subscore and total score approach 1, the reliability of the residual $S_{X \cdot Z}$ approaches 1, the coefficient of determination $\rho^2(S_V|S_Z)$ converges to $\rho^2(S_V, \tau_V)\rho^2(\tau_V, \tau_Z)$, and $\rho^2(S_V|S_Z, S_X)$ converges to

$$\rho^2(S_V|\tau_Z, \tau_X) = \rho^2(S_V, \tau_V)\{\rho^2(\tau_V, \tau_X) + [1 - \rho^2(\tau_X, \tau_Z)]\rho^2(\tau_V, \tau_X|\tau_Z)\},$$

where $\rho(\tau_V, \tau_X|\tau_Z)$ is the partial correlation of the true scores τ_X and τ_V given the true score τ_Z. Thus the incremental validity measure $\rho^2(S_V|S_{X \cdot Z})$ has a limit no greater than $\rho^2(S_V, \tau_V)[1 - \rho^2(\tau_X, \tau_Z)]$, so that high correlation of the true total score and the true subscore limits the incremental validity even when the reliability coefficients are high for both the subscore and the total score.

3 Examples

To illustrate results, the analysis in this section may be applied to the examples considered in Haberman (2008). In the first example (Tables 1, 2, and 3), subscores were examined from an SAT® I administration from 2002. In these tables, the SAT verbal examination is divided into the sections Verbal I, Verbal II, and Verbal III, while the SAT math examination is divided into the sections Math I, Math II, and Math III. Alternatively, the SAT verbal has sections for critical reading (CR), analogies (A), and sentence completion (SC), while the SAT math has sections for four-choice math multiple-choice (Math 4c), five-choice multiple choice (Math 5c), and student-produced math responses (Math S). Note that the SAT I examination of 2002 is substantially different from the current SAT Reasoning Test™, and reporting of these scores was confined to reports of raw scores on CR, A, and SC to examinees but not to institutions.

In each subscore in each table, the estimated reliability coefficient is sufficient so that relatively limited restriction on validity results is imposed. The smallest estimated reliability

Table 1

Estimated Reliability Coefficients of Subscores and Residual Subscores for SAT Verbal

Subscore	Subscore reliability	Residual subscore reliability
Verbal I	0.84	0.11
Verbal II	0.80	0.02
Verbal III	0.72	0.17
CR	0.84	0.24
A	0.74	0.16
SC	0.78	0.15

Note. CR = critical reading, A = analogies, SC = sentence completion.

Table 2

Estimated Reliability Coefficients of Subscores and Residual Subscores for SAT Math

Subscore	Subscore reliability	Residual subscore reliability
Math I	0.87	0.08
Math II	0.83	0.10
Math III	0.64	0.08
Math 4c	0.72	0.08
Math 5c	0.89	0.06
Math S	0.73	0.12

Note. Math 4c = four-choice math multiple-choice, Math 5c = five-choice math multiple-choice, Math S = student-produced math responses.

Table 3

Estimated Reliability Coefficients of Subscores and Residual Subscores for SAT Total

Subscore	Subscore reliability	Residual subscore reliability
Verbal	0.91	0.72
Math	0.92	0.72

coefficient is for Math III as a subscore of math. The coefficient of 0.64 for this case just implies that 0.64 is at least as large as the coefficient of determination for prediction of the criterion score by the Math III score. On the other hand, far stronger restrictions on incremental validity are present when subscores of verbal or math are considered. The extreme case is Verbal II, where the estimated reliability of the residual of 0.02 severely restricts incremental validity. The coefficient of determination for prediction of the criterion score by the Verbal II and the total verbal scores cannot by more than 0.02 greater than the coefficient of determination for prediction of the criterion score by the total verbal score. It is notable that no math subscore offers much possibility in terms of incremental validity, for the maximum reliability of a residual subscore is 0.12, a value achieved for Math S, the student-produced responses. Even here, the potential incremental improvement in prediction of a criterion score is quite limited. The coefficient of determination from use of the total math score to predict the criterion score cannot be more than 0.12 less than the coefficient of determination from prediction of the criterion score by both the total math score and the Math S subscore. In practice, further limits can be expected because the reliability of the criterion score is typically someone less than 1 and the correlation of the true residual score and the criterion score may well be somewhat less than 1.

The highest possibility among the subscores of the verbal and math tests is the critical reasoning portion of the verbal test, where the upper bound for incremental validity is 0.24. More generally, analogies, sentence completion, and Verbal III from the verbal test have higher potential for incremental validity than any subscores from the math test.

The situation is quite different when verbal and math are considered to be subscores of a total SAT score. Both subscores have quite high reliability, and the restriction on incremental validity is of little consequence, for the reliability of the residual subscores is 0.72.

For a second example, consider the Praxis™ examination results in Haberman (2008). Here subscores are present for English language arts (E), mathematics (M), citizenship and social science (C), and science (S). Results are summarized in Table 4. The subscore reliability coefficients, all of which are at least 0.68, provide only modest restrictions on the correlation of a validity criterion with the subscore. The restriction on incremental validity is somewhat less for English language arts and mathematics than for citizenship and social science and for science, for the residual subscore reliability coefficients have estimates of 0.25 and 0.29 for science and for citizenship and social science, respectively, while the corresponding reliability coefficients

Table 4
Estimated Reliability Coefficients of Subscores and Residual Subscores for Praxis Data

Subscore	Subscore reliability	Residual subscore reliability
E	0.73	0.43
M	0.79	0.48
C	0.68	0.29
S	0.69	0.25

Note. C = citizenship and social science, E = English language arts, M = mathematics, S = science.

for English language arts and mathematics are 0.43 and 0.49, respectively. Thus at least some possibility for appreciable incremental validity exists for all subscores.

4 Conclusions

In terms of validity, results in this report indicate that subscores have limited potential value unless they have some level of reliability and unless the true subscores are not very highly correlated with the true total score. Adequate reliability of the subscore is required for any possible validity result. The subscore cannot be highly correlated with a criterion score unless the subscore has a high reliability coefficient.

Even for a subscore with high reliability, potential validity results are limited when the correlation of the true subscore τ_X and the true remainder score τ_Y is quite high. This situation corresponds to a high correlation of the true subscore τ_X and the true total score $\tau_Z = \tau_X + \tau_Y$.

The limits on validity can have modest impact even if the correlation $\rho(\tau_X, \tau_Z)$ indicates a strong relationship between the true scores τ_X and τ_Z. This situation holds for the SAT math and verbal scales in Table 3, for the estimated value of $\rho(\tau_X, \tau_Z)$ exceeds 0.9 for these cases (Haberman, 2008).

In practice, limits on incremental validity should be examined to verify that subscores have any realistic possibility of usefulness. To be sure, some potential for utility does not ensure actual usefulness. On the other hand, negligible potential can eliminate any need for further study.

References

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

Holland, P. W., & Hoskens, M. (2003). Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test. *Psychometrika*, *68*, 123–149.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.