Educational Measurement
ISSUES AND PRACTICE

NCME national council on measurement in education

Educational Measurement: Issues and Practice
Summer 2014, Vol. 33, No. 2, pp. 29–40

Quantifying Error and Uncertainty Reductions in Scaling Functions: An ITEMS Module

Tim Moses, *Educational Testing Service*

This module describes and extends X-to-Y regression measures that have been proposed for use in the assessment of X-to-Y scaling and equating results. Measures are developed that are similar to those based on prediction error in regression analyses but that are directly suited to interests in scaling and equating evaluations. The regression and scaling function measures are compared in terms of their uncertainty reductions, error variances, and the contribution of true score and measurement error variances to the total error variances. The measures are also demonstrated as applied to an assessment of scaling results for a math test and a reading test. The results of these analyses illustrate the similarity of the regression and scaling measures for scaling situations when the tests have a correlation of at least .80, and also show the extent to which the measures can be adequate summaries of nonlinear regression and nonlinear scaling functions, and of heteroskedastic errors. After reading this module, readers will have a comprehensive understanding of the purposes, uses, and differences of regression and scaling functions.

Keywords: scaling, equating, concordance, regression, prediction error, scaling error

The application of observed score regression concepts in scaling and equating practice is a source of ongoing confusion in attempts to distinguish these methods. Regression and scaling/equating methods tend to be incorrectly described as the same approach to linking tests: "In linear equating, for example, scores on one test form are regressed on the other test form" (Embretson & Reise, 2000, p. 21). Even when the fundamental differences between regression and scaling or equating methods are specifically stated (Dorans & Holland, 2000; Kolen & Brennan, 2004; Livingston, 2004), regression concepts have been recommended for use in evaluating scaling functions (Dorans, 1999, 2004; Dorans & Walker, 2007). The purpose of this module is to review the recommended uses of regression error to evaluate scaling functions and to describe adaptations that produce error estimates more relevant for evaluating scaling and equating functions.

Using Regression Results to Evaluate Scaling Functions

The use of linear regression concepts to interpret scaling functions has a basis in concordance studies designed to link the scores of one test, X, to another test's scale, Y (Dorans, 1999, 2004; Dorans & Walker, 2007). In a concordance study, a link is established for the scales of two tests developed from different specifications, X and Y. The link of interest, referred to as a scaling function, produces a symmetric transformation of X such that the X-to-Y scale scores preserve Y's scale (i.e., Y's mean and standard deviation, μ_Y and σ_Y),

$$sc_y(X) = \mu_Y + \frac{\sigma_Y}{\sigma_X}(X - \mu_X). \tag{1}$$

In addition to producing a transformation that maintains the scales of X and Y, another intention of concordance studies is to use the concorded scores as estimates of actual Y scores for examinees with scores on X but not Y. The substitution of $sc_y(X)$ for an actual Y score is most accurate when X and Y are highly related. These issues have been heuristically described in terms of how the XY correlation, $\rho(X,Y)$, reduces the unexplained variability or "uncertainty" about a student's Y score. The percentage of Y's variance explained by the XY correlation, $\rho(X,Y)^2$, can be used in a measure showing how $\rho(X,Y)^2$ reduces Y's variance or uncertainty,

$$\frac{\sigma_Y - \sigma_Y\sqrt{1-\rho(X,Y)^2}}{\sigma_Y} = 1 - \sqrt{1-\rho(X,Y)^2}. \tag{2}$$

The right-hand side of Equation 2 has been described as a one minus the coefficient of alienation for regression function (McNemar, 1969), and also as an uncertainty reduction and a reduction in uncertainty (RiU) measure when predicting Y from X (Dorans, 1999, 2004; Dorans & Walker, 2007). Most of these prior descriptions have advocated the use of Equation 2 for making judgments about the quality of scaling functions. An example of a recommended judgment is that concordances should only be produced when tests X and Y have a correlation of at least .87 because uncertainty reductions from regression-based predictions should be at least 50% (Dorans, 1999, pp. 3, 5, and 15; Dorans & Walker, 2007, p. 185). One question prompted by the prior recommendations and judgments is whether Equation 2's correlation basis means that it is not directly relevant for evaluating X-to-Y scaling functions.

Adapting Regression Concepts to Produce More Direct Evaluations of Scaling Functions

The left hand side of Equation 2 is a relatively general expression that is used in this module to make Equation 2 directly relevant for the scaling functions defined in Equation 1. To develop a version of Equation 2 that indicates how much of Y's uncertainty is reduced by using an X-to-Y scaling function, well known results about X-to-Y regression functions (Pedhazur, 1997) are first reviewed and described in terms of their role in Equation 2. The X-to-Y regression function is defined as

$$\hat{Y}|X = \mu_Y + \rho(X,Y)\frac{\sigma_Y}{\sigma_X}(X - \mu_X). \tag{3}$$

The regression function's prediction error can be expressed as the variance of

$$Y - \hat{Y}|X = Y - \mu_Y - \rho(X,Y)\frac{\sigma_Y}{\sigma_X}(X - \mu_X), \tag{4}$$

which can be estimated as

$$\begin{aligned} &\sigma^2\left(Y - \hat{Y}|X\right) \\ &\quad = E\left\{Y - \hat{Y}|X - E\left[Y - \hat{Y}|X\right]\right\}^2, \\ &\quad = E\left[Y - \hat{Y}|X\right]^2, \\ &\quad = \sigma_Y^2\left[1 - \rho(X,Y)^2\right], \end{aligned} \tag{5}$$

where the Es denote expectations taken over the X and Y distributions. The square root of Equation 5's result, $\sigma_Y\sqrt{1-\rho(X,Y)^2}$, is the regression function's prediction error expressed in Y's standard deviation units. Equation 5's result is used in the left hand side of Equation 2's measure of uncertainty reduction.

A measure of uncertainty reduction that is directly relevant for an X-to-Y scaling function, $sc_y(X)$, can be produced by using adaptations of Equations 4 and 5 in the left-hand side of Equation 2. Similar to Equation 4's deviation from a regression function, the deviation of Y from Equation 1's scaling function can be obtained as

$$Y - sc_y(X) = Y - \mu_Y - \frac{\sigma_Y}{\sigma_X}(X - \mu_X). \tag{6}$$

Equation 5's approach can be applied to obtain the variance of Equation 6:

$$\begin{aligned} &\sigma^2(Y - sc_y(X)) \\ &\quad = E\{Y - sc_y(X) - E[Y - sc_y(X)]\}^2. \\ &\quad = E[Y - sc_y(X)]^2, \\ &\quad = 2\sigma_Y^2\,[1 - \rho(X,Y)]. \end{aligned} \tag{7}$$

Finally, the square root of Equation 7's result can be used with Equation 2 to obtain the proportion of Y's uncertainty that is reduced using an X-to-Y scaling function,

$$\frac{\sigma_Y - \sigma_Y\sqrt{2\,[1-\rho(X,Y)]}}{\sigma_Y} = 1 - \sqrt{2\,[1-\rho(X,Y)]}. \tag{8}$$

Equation 8's measure of uncertainty reduction based on Equation 7's error variance for scaling functions is somewhat different than Equation 2's measure based on a regression function's prediction error variance. In contrast to Equation 2's uncertainty reduction measure that suggests greater uncertainty reductions for XY correlations that are larger in absolute value, Equation 8's uncertainty reduction measure is reduced only when the XY correlation is greater than .5. Equation 8's values become small and even negative when the XY correlation is less than .5 because for correlations less than .5, Equation 7's scaling error variance ultimately reflects increased uncertainty about Y (i.e., scaling error variance that increases to a value greater than Y's variance). Negative XY correlations are especially problematic for Equation 7's scaling error variance because, unlike regression functions (Equation 3), scaling functions (Equation 1) always reflect positive XY relationships. The increased uncertainty about Y that results from computing scaling functions for tests that are not highly correlated is consistent with longstanding beliefs that equating and scaling functions for dissimilar tests are suspect (Kolen & Brennan, 2004, p. 129), should be discouraged (Angoff, 1954, p. 11), or should be replaced with regression functions (Dorans, 1999, p. 15). The possibility that Equation 8 can produce negative values that indicate increased uncertainty is a reflection of Equation 7's error variance, which makes Equation 8 similar to other measures of percent change that are able to take positive and negative values (e.g., percent changes in prices, home values, and retirement accounts).

Comparisons of Regression and Scaling Functions' Prediction Errors and Uncertainty Reductions

The implications of scaling error (Equation 7) and uncertainty reductions based on scaling functions (Equation 8) can be informatively described in a series of comparisons with the corresponding regression quantities (Equations 2 and 5). The next sections of this module provide comparisons of the scaling and regression error quantities, first for a hypothetical range of XY correlations, then in terms of how the error quantities are made up of true score variance and error variance, and finally as applied to empirical data from scaling studies for a math test and a reading test.

Comparisons of Uncertainty Reductions for a Range of XY Correlations

To gain insight into how uncertainty reductions based on regression functions (Equation 2) compare with those based on scaling functions (Equation 8), Table 1 and Figure 1 compare the two uncertainty reductions across a range of XY correlations. The results in Table 1 and Figure 1 show that, except for XY correlations of one, the uncertainty reductions are greater when based on regression functions. This result is unsurprising given that regression functions are constructed to minimize error (i.e., produce the minimum possible uncertainty in Y given X; Pedhazur, 1997). For XY correlations greater than .80, the two uncertainty reductions are very similar, and the uncertainty reductions based on scaling functions are only 1%–3% less than those based on regression functions. When XY correlations are less than or equal to .50, Table 1 and Figure 1 show that the uncertainty reduction values based on regression functions are equal for negative and positive correlations with the same absolute value whereas uncertainty reduction values based on scaling functions are much smaller than those based on regression functions, and assume values that decrease to –100%.

Table 1. Uncertainty Reductions From Regression and Scaling Functions for a Range of *XY* Correlations

ρ(*X*, *Y*)	% Uncertainty Reduction From a Regression Function (Equation 2) (%)	% Uncertainty Reduction From a Scaling Function (Equation 8) (%)
1.00	100	100
.95	69	68
.90	56	55
.87	50	48
.80	40	37
.75	34	29
.70	29	23
.65	24	16
.60	20	11
.55	16	5
.50	13	0
.45	11	–5
.40	8	–10
.35	6	–14
.30	5	–18
.25	3	–22
.20	2	–26
.15	1	–30
.10	1	–34
.05	0	–38
.00	0	–41
–.10	1	–48
–.20	2	–55
–.30	5	–61
–.40	8	–67
–.50	13	–73
–.60	20	–79
–.70	29	–84
–.80	40	–90
–.90	56	–95
–1.00	100	–100

Decomposing Regression and Scaling Error into Parts Due to True Score Variance and Measurement Error Variance

A useful way to evaluate and compare regression and scaling error is with respect to the parts of these functions' total error variances attributable to true score variance and to measurement error variance. Prior discussions have described the general correspondences of correlations, prediction errors, reliabilities, standard errors of measurement, and the strength of scaling functions (Dorans & Walker, 2007). The relationships of reliabilities, standard errors of measurement, regression error variance, and scaling error variance can be described more specifically when classical test theory assumptions about X and Y are integrated into Equations 1–8. From classical test theory assumptions the observed XY covariance matrix reflected in Equations 4 and 7 has a decomposition into matrices of true score variance and measurement error variance based on the reliabilities of X and Y (Lord & Novick, 1968),

$$\begin{aligned}\sum\nolimits_{XY} &= \sum\nolimits_{TX,TY} + \sum\nolimits_{\varepsilon X,\varepsilon Y} \\ &= \begin{bmatrix}\sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2\end{bmatrix} = \begin{bmatrix}\mathrm{rel}_X\sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \mathrm{rel}_Y\sigma_Y^2\end{bmatrix} \\ &\quad + \begin{bmatrix}(1-\mathrm{rel}_X)\sigma_X^2 & 0 \\ 0 & (1-\mathrm{rel}_Y)\sigma_Y^2\end{bmatrix}.\end{aligned} \tag{9}$$

The regression function's prediction error variance in Equation 5 can be similarly decomposed into parts attributable to true score variance and measurement error variance (Moses, 2012; Yen & Lu, in preparation),

$$\sigma^2\left(Y-\hat{Y}|X\right) = \sigma^2\left(Y-\hat{Y}|X\right)_{TX,TY} + \sigma^2\left(Y-\hat{Y}|X\right)_{\varepsilon X,\varepsilon Y}. \tag{10}$$

There are several ways in which Equation 10 can be viewed as an analogue to Equation 9:

- Equation 9's variance-covariance matrix of the observed X and Y scores, $\sum_{XY}$, is analogous to Equation 10's prediction error variance based on the observed X and Y scores, $\sigma^2(Y-\hat{Y}|X)$.
- Equation 9's variance-covariance matrix of X and Y's true scores, $\sum_{TX,TY}$, is analogous to Equation 10's prediction error variance based on of X and Y's true scores, $\sigma^2(Y-\hat{Y}|X)_{TX,TY}$.
- Equation 9's variance-covariance matrix of X and Y's measurement errors, $\sum_{\varepsilon X,\varepsilon Y}$, is analogous to Equation 10's prediction error variance based on of X and Y's measurement errors, $\sigma^2(Y-\hat{Y}|X)_{\varepsilon X,\varepsilon Y}$.
- Equation 9's decomposition of the variance-covariance matrix of the observed X and Y scores into the sum of variance-covariance matrices of X and Y's true scores and measurement errors is analogous to Equation 10's decomposition of prediction error variance based on observed X and Y scores into the sum of parts due to X and Y's true scores and measurement errors.

Further decompositions of the true score variance and error variance parts of Equation 10's prediction error variance are possible. The part of prediction error variance due to measurement error variance in Equation 10 can itself be decomposed into the parts due to the measurement error variances of X and of Y.

$$\sigma^2\left(Y-\hat{Y}|X\right)_{\varepsilon X,\varepsilon Y} = \sigma^2\left(Y-\hat{Y}|X\right)_{\varepsilon X} + \sigma^2\left(Y-\hat{Y}|X\right)_{\varepsilon Y}. \tag{11}$$

The part of prediction error variance due to true score variance in Equation 10 can itself be decomposed into a sum containing the true score variance from the true score regression and the variance of the difference between the observed and true score regressions:

$$\sigma^2\left(Y-\hat{Y}|X\right)_{TX,TY} = \sigma^2\left(TY-T\hat{Y}|TX\right)_{TX,TY} + \sigma^2\left(\hat{Y}|X,TX-T\hat{Y}|TX\right)_{TX,TY}. \tag{12}$$

Formulas for the parts of the regression function's prediction error variance in Equations 10–12 are summarized in Table 2.

Similar to Equations 10, Equation 7's scaling error variance can be decomposed as

$$\sigma^2\left[Y-sc_y(X)\right] = \sigma^2\left[Y-sc_y(X)\right]_{TX,TY} + \sigma^2\left[Y-sc_y(X)\right]_{\varepsilon X,\varepsilon Y}. \tag{13}$$

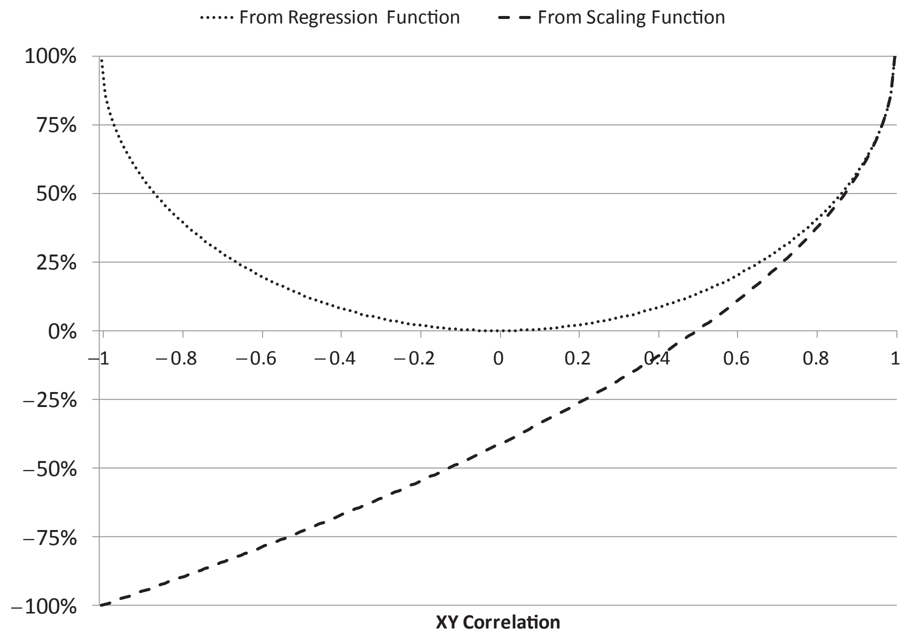

FIGURE 1. Uncertainty reductions from regression and scaling functions for a range of *XY* correlations.

Table 2. Comparison of Regression and Scaling Error Variances Attributable to *X* and *Y*'s True Score and Measurement Error Variance

	Regression/Scaling Error Variances and Their Parts	
Sources of Error	**From a Regression Function**	**From a Scaling Function**
Total Error Variance For the Observed Function	$\sigma_Y^2\left[1-\rho(X,Y)^2\right]$	$2\sigma_Y^2\left[1-\rho(X,Y)\right]$
Total True Score Variance From the Observed Function[a]	$\sigma_Y^2\left[\text{rel}_Y+\rho(X,Y)^2(\text{rel}_X-2)\right]$	$2\sigma_Y^2\left[\frac{\text{rel}_X+\text{rel}_Y}{2}-\rho(X,Y)\right]$
True Score Variance From the True Score Function	$\sigma_Y^2\left[\text{rel}_Y-\frac{\rho(X,Y)^2}{\text{rel}_X}\right]$	$2\sigma_Y^2\left[\text{rel}_Y-\frac{\sqrt{\text{rel}_Y}}{\sqrt{\text{rel}_X}}\rho(X,Y)\right]$
Variance of the Difference Between True and Observed Functions	$\text{rel}_X\sigma_X^2\left[\frac{\sigma_{XY}}{\sigma_X^2}-\frac{\sigma_{XY}}{\text{rel}_X\sigma_X^2}\right]^2$	$\sigma_Y^2\left[\text{rel}_X+\text{rel}_Y-2\sqrt{\text{rel}_X}\sqrt{\text{rel}_Y}\right]$
Total Measurement Error Variance From the Observed Function due to *X* and *Y*'s Measurement Errors	$\sigma_Y^2\left[(1-\text{rel}_Y)+\rho(X,Y)^2(1-\text{rel}_X)\right]$	$2\sigma_Y^2\left[1-\frac{\text{rel}_X+\text{rel}_Y}{2}\right]$
Measurement Error Variance due to *X*'s Measurement Error	$\sigma_Y^2\rho(X,Y)^2(1-\text{rel}_X)$	$\sigma_Y^2\left(1-\text{rel}_X\right)$
Measurement Error Variance due to *Y*'s Measurement Error	$\sigma_Y^2\left(1-\text{rel}_Y\right)$	$\sigma_Y^2\left(1-\text{rel}_Y\right)$

[a]For regression functions, total true score variance is a sum of the true score variance from the true regression function and the variance of the difference between the true and observed regression functions. This result is not obtained for scaling functions because for scaling functions the two true score variance parts have a nonzero covariance, $\text{rel}_Y\sigma_Y^2\left(\frac{\sqrt{\text{rel}_X}}{\sqrt{\text{rel}_Y}}-1+\frac{\rho(X,Y)}{\sqrt{\text{rel}_Y}\sqrt{\text{rel}_X}}-\frac{\rho(X,Y)}{\text{rel}_Y}\right)$.

The part of scaling error variance due to measurement error variance in Equation 13 can itself be decomposed into the parts due to the measurement error variances of X and of Y:

$$\sigma^2[Y-sc_y(X)]_{\varepsilon X,\varepsilon Y}=\sigma^2[Y-sc_y(X)]_{\varepsilon X}+\sigma^2[Y-sc_y(X)]_{\varepsilon Y}. \quad (14)$$

The part of scaling error variance due to true score variance in Equation 13 can itself be decomposed into a sum containing the true score variance from the true score scaling function and the variance of the difference between the observed and true score scaling functions:

$$\sigma^2\left[Y-sc_y(X)\right]_{TX,TY}=\sigma^2\left[TY-sc_{TY}(TX)\right]_{TX,TY}+\sigma^2\left[sc_y(X,TX)-sc_{TY}(TX)\right]_{TX,TY}. \quad (15)$$

Formulas for the parts of the scaling function's error variance in Equations 13–15 are summarized in Table 2.

Table 2 suggests that the major differences in the parts of regression and scaling error are whether the *XY* correlation affects the contribution of X's true score variance and measurement error variance. That is, in contrast to scaling error, X's contribution to the regression error's true score variance and measurement error variance depend on the XY correlation. For regression functions, X makes a relatively small percentage contribution to measurement error variance and a relatively large percentage contribution to the variance reflecting the difference between the observed and true regression functions.

Table 3 shows the regression and scaling uncertainty reductions and several additional percentages constructed from Table 2 for situations where $\sigma_X^2 = \sigma_Y^2 = 1$. The additional percentages in Table 3 were computed by taking the parts of regression and scaling error variances shown in Table 2 and dividing them by the total regression or scaling error variances. These percentages were formed from the variances rather than the standard deviations used in the uncertainty reduction ratios because ratios of the parts of Table 2's error variances directly sum to 100%.

Beginning with the bottom row of Table 3, where the XY correlation, X reliability and Y reliability are all .7, the percentage in Y's uncertainty reduction is 29% for a regression function and 23% for a scaling function. Table 3 also shows that these functions' error variances have different compositions, where 12%, 29%, and 59% of the regression function's error variance is made up of true versus observed variance, X's measurement error variance and Y's measurement error variance, and 0%, 50%, and 50% of the scaling function's error variance is made up of true versus observed variance, X's measurement error variance and Y's measurement error variance. The percentage differences are a direct reflection of the XY correlation being incorporated into the X-to-Y regression function but not the X-to-Y scaling function. That is, the regression function's use of the XY correlation results in a larger uncertainty reduction, a larger percentage contribution of true versus observed function variance and a smaller percentage contribution of X's measurement error variance. The other results in Table 3 are similar to those in the bottom row, in that with the regression function the uncertainty reduction is larger, the percentage contribution of true versus observed function variance is also larger, and the percentage contribution of X's measurement error variance is smaller.

Regression Error, Scaling Error, and Uncertainty Reductions in Scaling Studies

This module's results can be further considered by applying them to assess the results of actual scaling studies conducted to link two scales (A and B) of a math test and a reading test. The descriptive statistics for the tests' scale scores are shown in Table 4. In the original scaling studies for these data, equipercentile methods (i.e., nonlinear versions of Equation 1's scaling functions, $e_y(X)$) were used to link the reading and math tests' A scales to the B scales. These original studies prompted questions about how to most appropriately convey the equipercentile scaling functions' inaccuracies. Measures of error due to sampling variability and subgroup variability typically reported in scaling studies were not considered desirable because these measures do not directly indicate the variability of examinees' scale B scores from the A-to-B scaling functions.

This module's results can be applied to address questions about variability from the A-to-B scaling functions for Table 4's math and reading tests. The uncertainty reductions and variance percentages based on the reading and math measures' correlations, reliabilities, estimated true score variances and estimated measurement error variances are shown in Table 5. For the reading test, A-to-B regression and scaling functions reduce the uncertainty about examinees' scale B scores by 43% and 41%. For the math test, the uncertainty reductions for the regression and scaling functions are 39% and 35%. Because the reliabilities of the A and B scale scores are greater than .9 and the correlations of the A and B scale scores are about .8, more than 50% of the error variances of the regression and scaling functions is attributable to true score variance and less than 50% to measurement error variance. Table 6 presents the regression and scaling errors in standard deviation units (the square roots of Equations 5 and 7), and their parts attributable to true score variance and measurement error variance (the square roots of the quantities in Table 2). Table 6 shows that the total errors for A-to-B linear regression and scaling functions are approximately six standard deviations on the tests' B scales (i.e., the fourth and fifth columns), reflective of true score variances of more than four standard deviations (i.e., the sixth through ninth columns) and measurement error variances of about four standard deviations (i.e., the 10th through 13th columns). Altogether these results suggest that the total variances of regression and scaling functions for the A and B scales of the math and reading tests are due more to the tests' measurement differences (i.e., true score variance) than to the tests' measurement errors (i.e., measurement error and unreliability).

Conditional and Overall Regression and Scaling Error in the Scaling Results

The results of Tables 5 and 6 can be described as averages or summaries of the variances of the A-to-B linear regression and linear scaling functions. Tables 5 and 6 prompt two additional questions. One question is about the adequacy of the error variance estimates of the linear regression and linear scaling functions results for approximating the error variances of nonlinear regression and nonlinear scaling functions that fit the data more closely than the linear functions. This question can be addressed by computing regression and scaling error variances using nonlinear regression functions (i.e., the conditional means of the scale B scores given scale A scores, $\mu_{B|A}$) and using nonlinear scaling functions (i.e., the equipercentile functions actually produced in the scaling studies, $e_B(A)$) in Equations 5 and 7. The error variances for the nonlinear regression functions are $\sqrt{E[B-\mu_{B|A}]^2} = 5.740$ for the reading test and 5.733 for the math test, values which are close to, but slightly smaller than the standard deviations from the linear regression functions (5.80 and 5.85). The error variances from the equipercentile functions are $\sqrt{E[B-e_B(A)]^2} = 6.016$ for the reading test and 6.085 for the math test, values which are close to, but slightly smaller than the standard deviations from the linear regression functions (6.07 and 6.18). These results suggest that the error variance estimates of the linear regression and linear scaling functions slightly overestimate the error variances of the nonlinear functions.

A second question is about the adequacy of the Table 5 and 6 error estimates of the linear regression and linear scaling functions for approximating conditional errors that are not necessarily constant. The actual conditional regression error for a nonlinear regression function can be obtained as the standard deviations around the A-to-B conditional means

Table 3. Comparison of Regression and Scaling Functions' Uncertainty Reductions, and Decompositions of their Error Variances Into %'s Attributable to X and Y's True Score and Measurement Error Variance

			% Uncertainty Reduction (%)		% of Regression/ Scaling Error Due to True Score Variances From the True Score Functions (%)		% of Regression/ Scaling Error Due to Variances of the Difference Between the True and Observed Functions (%)		% of Regression/ Scaling Error Due to Measurement Error Variances (*X*) (%)		% of Regression/ Scaling Error Due to Measurement Error Variances (*Y*) (%)	
$\rho(X, Y)$	rel_X	rel_Y	Regression	Scaling	Regression	Scaling	Regression	Scaling	Regression	Scaling	Regression	Scaling
.9	1.0	1.0	56	55	100	100	0	0	0	0	0	0
.9	1.0	.9	56	55	47	46	0	1	0	0	53	50
.9	.9	1.0	56	55	53	51	5	1	43	50	0	0
.9	.9	.9	56	55	0	0	5	0	43	50	53	50
.8	1.0	1.0	40	37	100	100	0	0	0	0	0	0
.8	1.0	.9	40	37	72	71	0	1	0	0	28	25
.8	1.0	.8	40	37	44	42	0	3	0	0	56	50
.8	.9	1.0	40	37	80	78	2	1	18	25	0	0
.8	.9	.9	40	37	52	50	2	0	18	25	28	25
.8	.9	.8	40	37	25	23	2	1	18	25	56	50
.8	.8	1.0	40	37	56	53	9	3	36	50	0	0
.8	.8	.9	40	37	28	26	9	1	36	50	28	25
.8	.8	.8	40	37	0	0	9	0	36	50	56	50
.7	.9	.9	29	23	70	67	1	0	10	17	20	17
.7	.9	.8	29	23	50	47	1	0	10	17	39	33
.7	.9	.7	29	23	31	28	1	2	10	17	59	50
.7	.8	.9	29	23	56	53	5	0	19	33	20	17
.7	.8	.8	29	23	37	33	5	0	19	33	39	33
.7	.8	.7	29	23	17	15	5	1	19	33	59	50
.7	.7	.9	29	23	39	35	12	2	29	50	20	17
.7	.7	.8	29	23	20	17	12	1	29	50	39	33
.7	.7	.7	29	23	0	0	12	0	29	50	59	50

Table 4. Descriptive Statistics for the Data From Scaling Studies of a Reading and Math Test

	Min Observed	Max Observed	*N*	Mean	Std. Dev.	Skew	Reliabilities	Correlations of Tests' Scale A and B Scores
Reading Test: Scale A	21	85	1,194	50.69	11.62	.089	.93	.824
Reading Test: Scale B	3	50	1,194	27.43	10.22	–.012	.91	
Math Test: Scale A	16	83	1,157	62.50	12.45	–.731	.92	.791
Math Test: Scale B	5	49	1,157	30.12	9.55	–.077	.90	

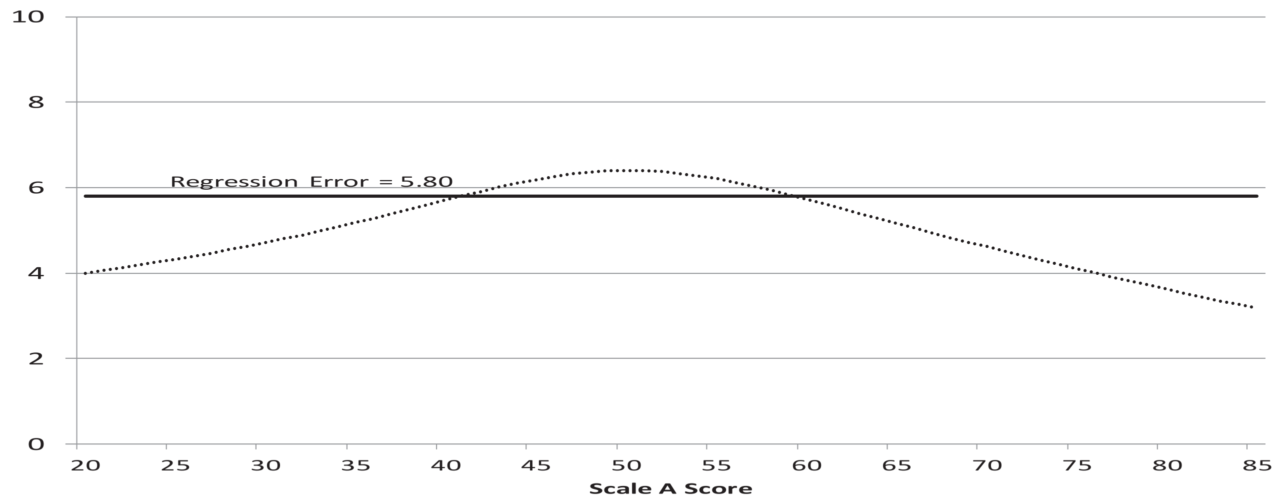

FIGURE 2. Conditional standard deviations from the conditional means, Reading test.

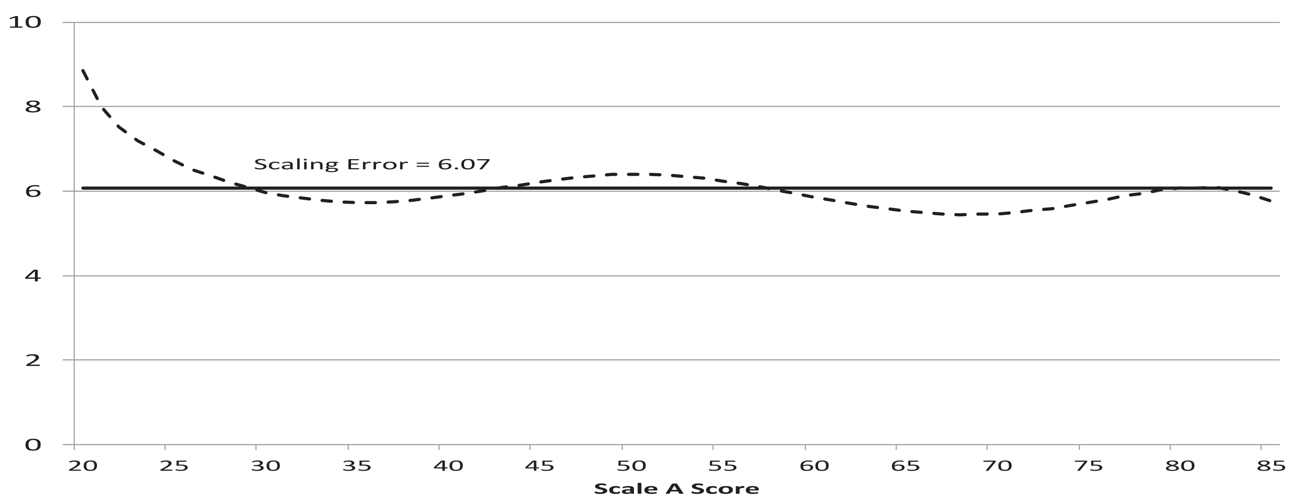

FIGURE 3. Conditional standard deviations from the equipercentile scaling function, Reading test.

computed at each A score, $\sqrt{E[B-\mu_{B|A}]^2|A}$. The actual conditional scaling error for a nonlinear scaling function can be obtained as the standard deviations around the A-to-B equipercentile scaling functions computed at each A score, $\sqrt{E[B-e_B(A)]^2|A}$. Figures 2–7 illustrate the conditional error estimates and the estimates of total regression error from the linear regression function (the square root of Equation 5) and total scaling error from the linear scaling function (the square root of Equation 7). Figures 2–7 verify that the conditional standard deviations are not perfectly constant across the scale A scores, but may be interpretable as generally reflective of total regression and scaling error. The conditional standard deviations are nearest to the total regression and scaling error estimates around the means of the scale A scores. Similar to how the regression errors are smaller than the scaling errors (5.80 vs. 6.07 for reading, 5.85 vs. 6.18 for math), Figures 4 and 7 show that the conditional standard deviations around the means are smaller than the conditional standard deviations around the equipercentile scaling functions, with the differences being largest at the highest and lowest scale

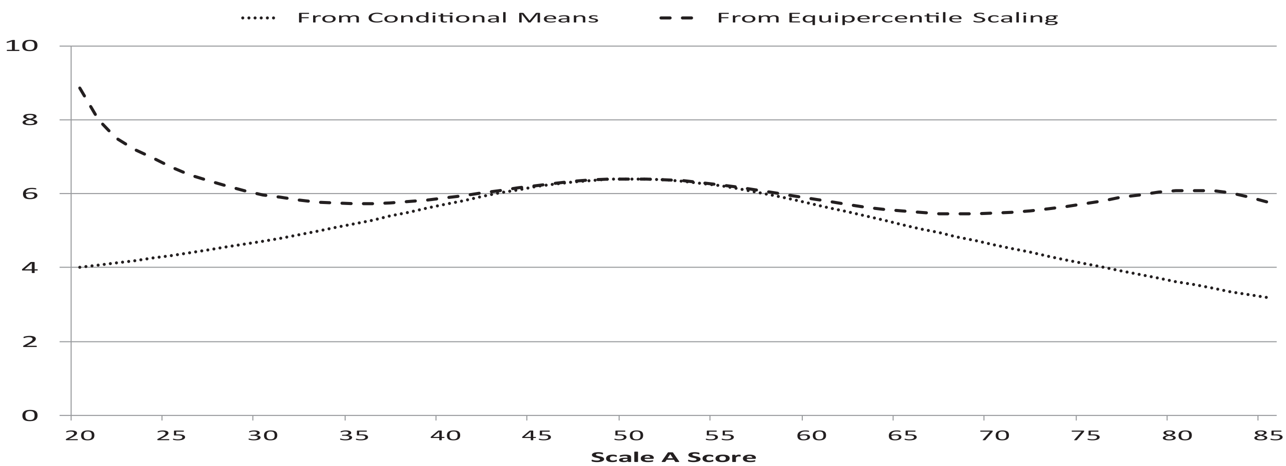

FIGURE 4. Conditional standard deviations from the conditional means and the equipercentile scaling function, Reading test.

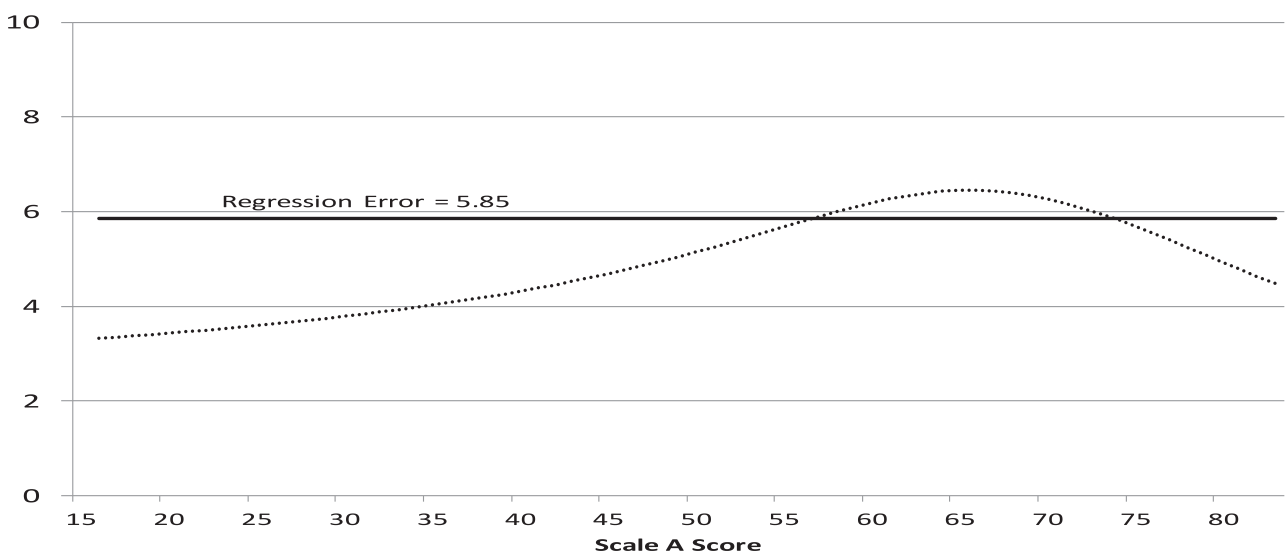

FIGURE 5. Conditional standard deviations from the conditional means, Math test.

A scores and negligibly smaller near the mean of the scale A scores.

Summary and Concluding Remarks

Prior discussions of regression and scaling functions have not always pointed out these functions' differences, implying either that the functions are equal (Embretson & Reise, 2000) or proposing that measures based on correlations and regression functions be used in their original form to evaluate scaling functions (Dorans, 1999, 2004; Dorans & Walker, 2007). These discussions do not completely represent psychometric practice where scaling functions may be preferred over regression functions because scaling functions are symmetric, directly maintain tests' scales, and avoid regression to the mean (Dorans & Holland, 2000; Kolen & Brennan, 2004; Livingston, 2004) but may also be used like regression functions to predict students' missing scores (i.e., concordance studies). The differences between regression and scaling functions would seem to make the use of regression-based measures inadequate for evaluating scaling functions, but prior discussions have not been specific about these inadequacies.

This module was intended to clarify prior discussions of regression and scaling functions. The measures that were developed to quantify prediction error from scaling functions (Equations 7 and 8) were useful for indicating the potential loss in prediction accuracy when using a scaling function rather than a regression function (Table 1), and for clarifying how tests' measurement characteristics affect the prediction errors of regression and scaling functions (Tables 2 and 3). More general developments of this module's measures are possible, such as those that utilize linear functions other than regression and scaling functions, different kinds of nonlinear regression and scaling functions, and error variance estimates for the equating and scaling functions computed from data collection designs where the test scores' correlation cannot be directly estimated (e.g., equivalent groups and anchor test designs).

Self-Test

1. What function would be preferred for the goal of predicting a test taker's score on another test as precisely as possible?

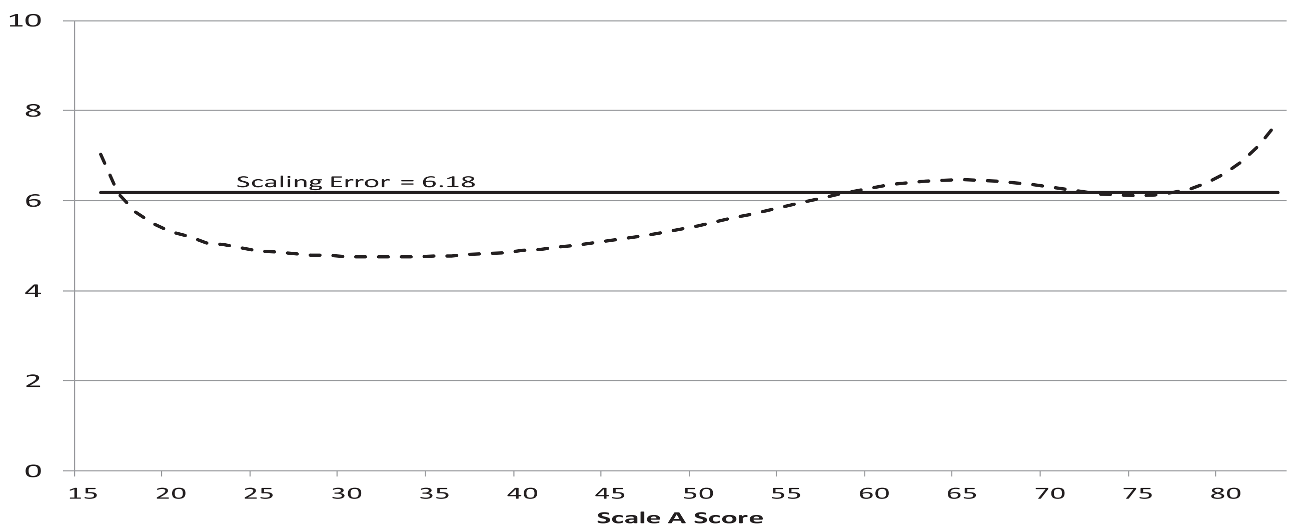

FIGURE 6. Conditional standard deviations from the equipercentile scaling function, Math test.

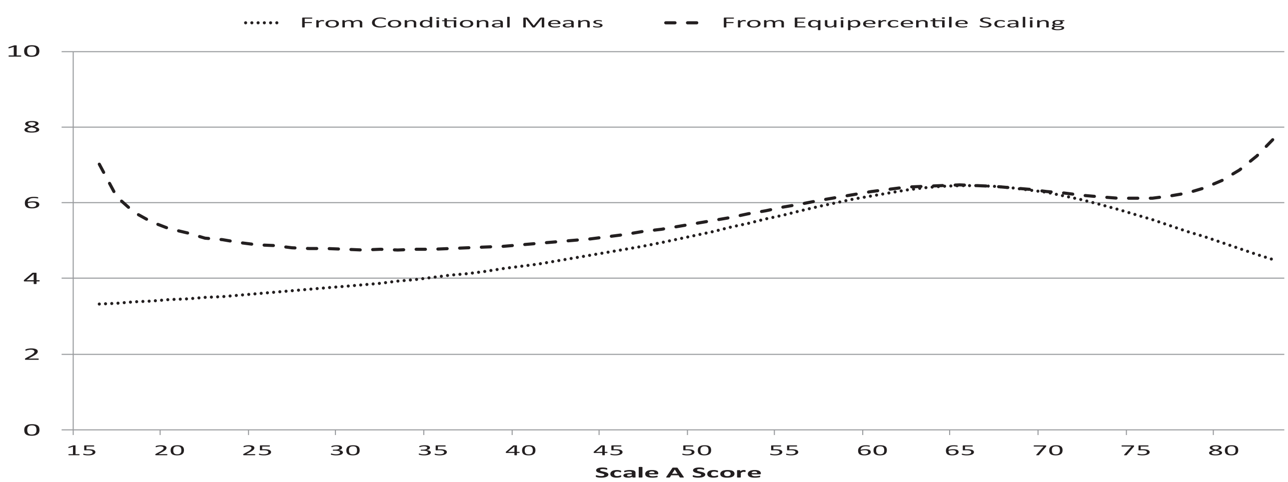

FIGURE 7. Conditional standard deviations from the conditional means and the equipercentile scaling function, Math test.

2. What function would be preferred for the goal of maintaining the scales of two tests?
3. What factor determines how well regression and scaling functions will simultaneously meet the goals of predicting an examinee's score on one test from another test and maintaining two tests' scales?
4. For two tests whose relationship is being studied, what measurement characteristics would influence the factor addressed in question three?
5. Consider the situation where a request has been made to use scaling methods to produce a concordance of two tests, X and Y, which have a correlation less than .5. Assume that the intended purpose of this concordance is to encourage test takers with a score on X but not Y to use the concordance table to estimate their score on Y from their score on X. What cautions would be warranted for the requester(s) and the potential test users?
6. Develop R computer code that will take two sets of test scores, along with an estimate of the reliability for each, and compute regression and scaling error results such as those described in this module.

Answers to Self-Test

1. Regression functions like Equation 3 are estimated based on minimizing prediction error.
2. Scaling functions like Equation 1 are estimated based on maintaining the tests' scales.
3. Regression and scaling functions are most similar when the tests' correlation approaches +1.
4. The strength of the observed correlation is influenced by tests' reliabilities and their measurement similarity (i.e., the tests' covariance or true score correlation).
5. It would be appropriate to warn the requesters and test users that the concordance table's predictions will not be very accurate and that more accurate predictions could be obtained from alternative methods. Table 1 shows that predictions based on regression methods would be more accurate (i.e., larger uncertainty reductions) than those based on scaling methods. Table 1 also indicates that for correlations less than 0.5, predictions from the scaling function will produce negative uncertainty reductions, implying that these predictions are actually less accurate than those based on ignoring the test user's

Table 5. Comparison of Regression and Scaling Functions' Uncertainty Reductions, and Decompositions of Their Error Variances Into %'s Attributable to Tests' A and B Scale Scores' True Score and Measurement Error Variance From the Scaling Studies

Correlations of Tests' Scale A and B Scores	rel_X Scale A Scores	rel_Y Scale B Scores	% Uncertainty Reduction (%)		% of Regression/ Scaling Error Due to True Score Variances From the True Score Functions (%)		% of Regression/ Scaling Error Due to Variances of the Difference Between the True and Observed Functions (%)		% of Regression/ Scaling Error Due to Measurement Error Variances (Scale A Scores) (%)		% of Regression/ Scaling Error Due to Measurement Error Variances (Scale B Scores) (%)	
			Regression	Scaling	Regression	Scaling	Regression	Scaling	Regression	Scaling	Regression	Scaling
Reading Test Results												
.824	.93	.91	43	41	56	54	1	0	15	20	28	26
Math Test Results												
.791	.92	.90	39	35	59	56	1	0	13	19	27	24

Table 6. Comparison of Regression and Scaling Functions' Errors and the Decompositions of the Errors Into Parts Attributable to Tests' A and B Scale Scores' True Score and Measurement Error Variance

Correlations of Tests' Scale A and B Scores	rel_X Scale A Scores	rel_Y Scale B Scores	Regression/ Scaling Error		Regression/Scaling Error Due to True Score Variances From the True Score Functions		Regression/Scaling Error Due to Variances of the Difference Between the True and Observed Functions		Regression/Scaling Error Due to Measurement Error Variances (Scale A Scores)		Regression/Scaling Error Due to Measurement Error Variances (Scale B Scores)	
			Regression	Scaling	Regression	Scaling	Regression	Scaling	Regression	Scaling	Regression	Scaling
Reading Test Results												
.824	.93	.91	5.80	6.07	4.35	4.46	.61	.11	2.23	2.70	3.07	3.07
Math Test Results												
.791	.92	.90	5.85	6.18	4.49	4.64	.63	.10	2.14	2.70	3.02	3.02

Note. All numbers are in standard deviation units of the Scale B scores from the scaling studies.

X score and simply using test *Y's* population mean as the predicted *Y* score.

6. R code and results are demonstrated for the Math tests described in Tables 4–6.

```
# Read in the data for the Math tests (assumes that the file directory is correctly specified).
ab=read.csv(file="math.csv")
a=ab[,1]
b=ab[,2]

# Create a function that produces the regression error results for two tests' data and reliabilities.
regressionerrorfunc=function(x,y,relx,rely) {
varp=function(z) sum((z-mean(z))^2/length(z))
regressionuncertaintyred=100*(1-sqrt(1-cor(x,y)^2))
cat("Uncertainty Reduction for the Linear Regression Function:\n")
print(regressionuncertaintyred)
regressionerrortotal=varp(y)*(1-cor(x,y)^2)
regressionerrortotaltruevar=varp(y)*(rely+cor(x,y)
^2*(relx-2))
regressionerrortruefromtrue=varp(y)*(rely-
cor(x,y)^2/relx)
regressionerrorobstruevar=relx*varp(x)*(cor(x,y)*sqrt
(varp(y))/sqrt(varp(x))-
cor(x,y)*sqrt(varp(y))/(relx*sqrt(varp(x))))^2
regressionerrortotalxyerror=varp(y)*((1-
rely)+cor(x,y)^2*(1-relx))

regressionerrorxerror=varp(y)*(cor(x,y)^2*(1-relx))
regressionerroryerror=varp(y)*(1-rely)
regressionerrors=data.frame(regressionerrortotal, regressionerrortotaltruevar, regressionerrortruefromtrue, regressionerrorobstruevar, regressionerrortotalxyerror, regressionerrorxerror, regressionerroryerror)
cat("\n")
cat("Results in Variance Units:\n")

print(regressionerrors)
regressionerrorsp=100*regressionerrors/
regressionerrortotal
cat("\n")
cat("Results in Percentage Units:\n")
print(regressionerrorsp)
regressionerrorssd=sqrt(regressionerrors)
cat("\n")
cat("Results in Standard Deviation Units:\n")
print(regressionerrorssd)
}

# Create a function that produces the scaling error results for two tests' data and reliabilities.
scalingerrorfunc=function(x,y,relx,rely) {
varp=function(z) sum((z-mean(z))^2/length(z))
scalinguncertaintyred=100*(1-sqrt(2*(1-cor(x,y))))
cat("Uncertainty Reduction for the Linear Scaling Function:\n")
print(scalinguncertaintyred)
scalingerrortotal=2*varp(y)*(1-cor(x,y))
scalingerrortotaltruevar=2*varp(y)*((relx+rely)/2-
cor(x,y))
scalingerrortruefromtrue=2*varp(y)*(rely-
sqrt(rely)/sqrt(relx)*cor(x,y))
scalingerrorobstruevar=varp(y)*(relx+rely-
2*sqrt(relx)*sqrt(rely))
scalingerrortotalxyerror=2*varp(y)*(1-(relx+rely)/2)
scalingerrorxerror=varp(y)*(1-relx)
scalingerroryerror=varp(y)*(1-rely)

scalingerrors=data.frame(scalingerrortotal,
scalingerrortotaltruevar,scalingerrortruefromtrue,
scalingerrorobstruevar, scalingerrortotalxyerror,scalingerrorxerror,scalingerroryerror)
cat("\n")
cat("Results in Variance Units:\n")
print(scalingerrors)
scalingerrorsp=100*scalingerrors/scalingerrortotal
cat("\n")
cat("Results in Percentage Units:\n")
print(scalingerrorsp)
scalingerrorssd=sqrt(scalingerrors)
cat("\n")
cat("Results in Standard Deviation Units:\n")
print(scalingerrorssd)
}

# Run the regression error function and obtain the results.
regressionerrorfunc(a,b,.92,.90)
Uncertainty Reduction for the Linear Regression Function:
[1] 38.76642

Results in Variance Units:
regressionerrortotal regressionerrortotaltruevar regressionerrortruefromtrue
34.19329 20.51401 20.11749
regressionerrorobstruevar regressionerrortotalxyerror regressionerrorxerror
0.3965197 13.67928 4.559977
regressionerroryerror
9.1193

Results in Percentage Units:
regressionerrortotal regressionerrortotaltruevar regressionerrortruefromtrue
100 59.99426 58.83462
regressionerrorobstruevar regressionerrortotalxyerror regressionerrorxerror
1.159642 40.00574 13.33588
regressionerroryerror
26.66986

Results in Standard Deviation Units:
regressionerrortotal regressionerrortotaltruevar regressionerrortruefromtrue
5.847503 4.529239 4.485252
regressionerrorobstruevar regressionerrortotalxyerror regressionerrorxerror
0.6296981 3.698551 2.13541
regressionerroryerror
3.019818

# Run the scaling error function and obtain the results.
scalingerrorfunc(a,b,.92,.90)
Uncertainty Reduction for the Linear Scaling Function:
[1] 35.2849

Results in Variance Units:
```

scalingerrortotal scalingerrortotaltruevar scalingerrortruefromtrue
38.19203 21.77729 21.52937
scalingerrorobstruevar scalingerrortotalxyerror scalingerrorxerror
0.01002151 16.41474 7.29544
scalingerroryerror
9.1193

Results in Percentage Units:
scalingerrortotal scalingerrortotaltruevar scalingerrortruefromtrue
100 57.02051 56.37137
scalingerrorobstruevar scalingerrortotalxyerror scalingerrorxerror
0.02623979 42.97949 19.10199
scalingerroryerror
23.87749

Results in Standard Deviation Units:
scalingerrortotal scalingerrortotaltruevar scalingerrortruefromtrue
6.17997 4.666615 4.639975
scalingerrorobstruevar scalingerrortotalxyerror scalingerrorxerror
0.1001075 4.051511 2.701007
scalingerroryerror
3.019818

Acknowledgments

The author thanks Neil Dorans, W. Holmes Finch, Marna Golub-Smith, and Rebecca Zwick for helpful comments.

References

Angoff, W. H. (1954). *Comparable scores on ACE and SAT* (ETS Research Memorandum No. 54–6). Princeton, NJ: Educational Testing Service.

Dorans, N. J. (1999). *Correspondences between ACT and SAT I scores* (College Board Report No. 99-1). New York, NY: The College Board.

Dorans, N. J. (2004). Equating, concordance and expectation. *Applied Psychological Measurement*, *28*, 227–246.

Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, *37*, 281–306.

Dorans, N. J., & Walker, M. E. (2007). Sizing up linkages. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 179–198). New York, NY: Springer.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Hillsdale, NJ: Lawrence Erlbaum.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer-Verlag.

Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ: Educational Testing Service.

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

McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York, NY: John Wiley.

Moses, T. (2012). Relationships of measurement error and prediction error in observed-score regression. *Journal of Educational Measurement*, *49*, 380–398.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research*. Fort Worth, TX: Wadsworth.

Yen, W. M., & Lu, Y. (in preparation). The standard error of measurement as it relates to growth scores. Unpublished manuscript.

