# Rivier University

Education Division

Specialist in Assessment of Intellectual Functioning (SAIF) Program

ED 656, 657, 658, & 659

Psy.D. in Counseling and School Psychology Program ED 810

John O. Willis, Ed.D., SAIF

***ACHIEVEMENT (OR OTHER) SCORES PREDICTED FROM ABILITY SCORES***

***(Regression towards mediocrity)***

Some slides stolen
from Professor
Ron Dumont, Ed.D., NCSP
Director: School of Psychology,
Fairleigh Dickinson University
(formerly Senior Lecturer,
Rivier College)

# ACHIEVEMENT SCORES PREDICTED FROM ABILITY SCORES

**Scores are standard scores with M = 100 s.d. = 15**

**Correlations between ability and achievement tests are .50, .70, .90, and 1.00.**

**z score for predicted achievement = correlation * z score for ability measure**

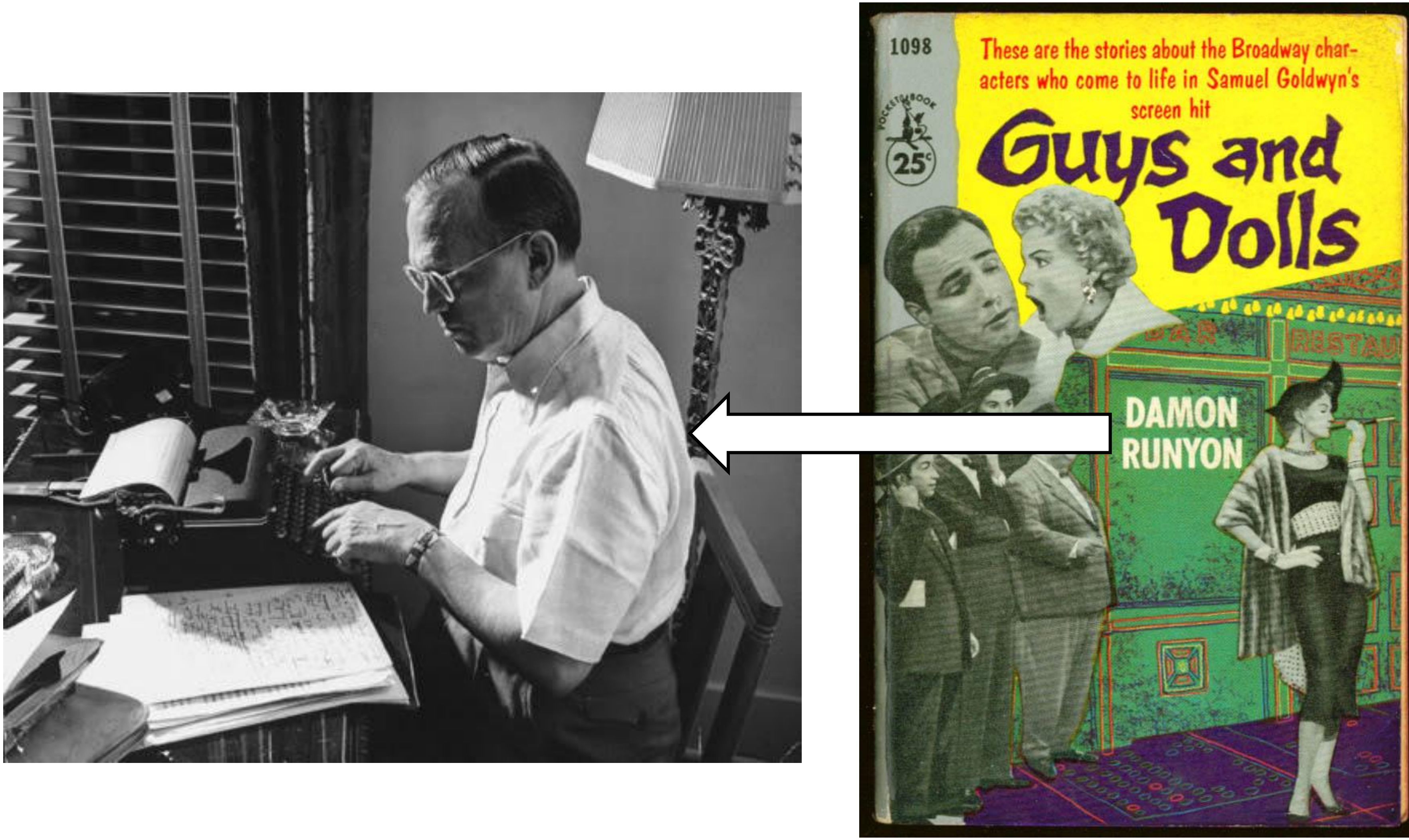
**z(predicted) = rxx * z(ability)**

These notes apply to any prediction of a score on one measure from a score on another, but the most common application in school assessments is, for some obscure reason, prediction of academic achievement from tests of cognitive ability.

I am not sure why one would want to predict reading fluency from a test of solving visual matrix puzzles, or any skill from any other, but this is how one would do it.

(If you want to know somebody's reading level, why not just use a reading test?)

I have heard that Damon Runyon insisted he never said, "The race is not always to the swift nor the battle to the strong, but that is the way to bet."


1098
These are the stories about the Broadway characters who come to life in Samuel Goldwyn's screen hit
25¢
Guys and Dolls
DAMON RUNYON

When predicting test scores without any information, the way to bet is the mean (50th percentile, standard score 100,[1] T score 50, etc.).

[1] As a shorthand, I use the term "standard score" to refer to a standard score with a mean of 100 and standard deviation of 15 unless I note otherwise.  All scores defined by their means and standard deviations (e.g., T scores, z-scores, scaled scores, *v*-scale scores, etc.) are standard scores, but I am confining the term to the standard scores used, for example, for Woodcock-Johnson tests, Wechsler index and IQ scores, and DAS clusters and GCA scores.

In a normal distribution, the mean (standard score 100) is not only the mean, but also the median (middle score) and mode (most frequent score), so that is indeed the way to bet if you have no information about the examinee.

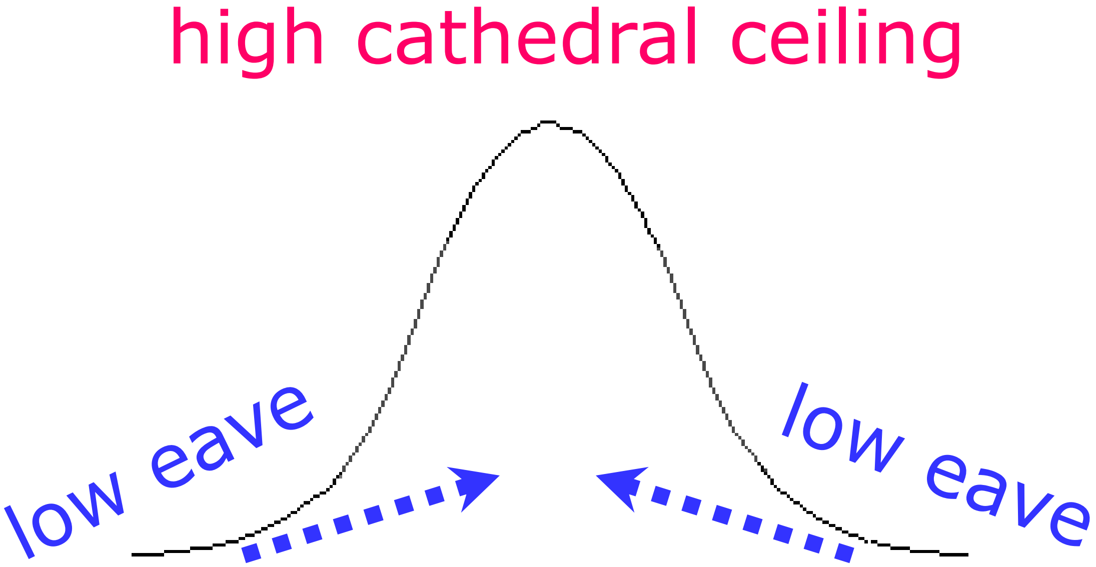
When you are trying to predict one score from another score, it is an actual, non-alternative fact of life that the predicted score will be closer to the mean (100) than is the score from which you make the prediction (the predictor).

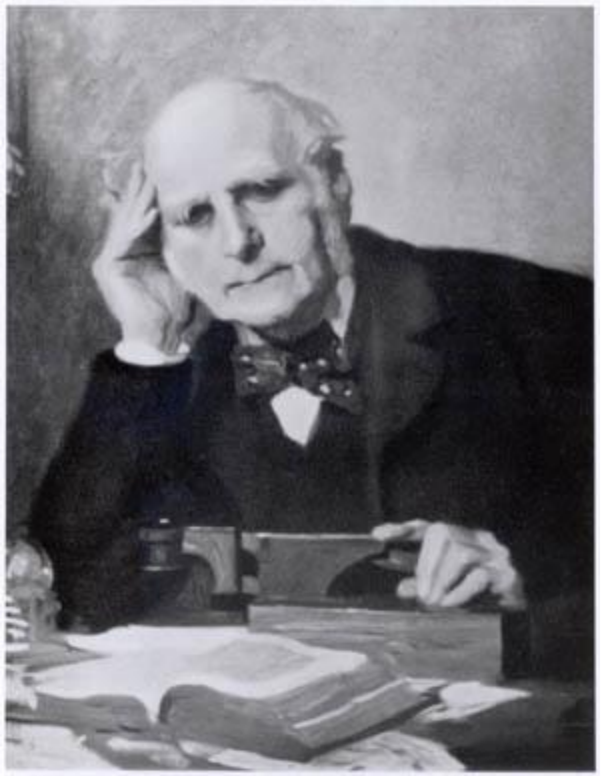
If you start with a score on one variable (the predictor, e.g., IQ) and then test another variable (the predicted, e.g., academic achievement or adaptive behavior), most (not all) folks will score closer to 100 on the predicted score than they did on the predictor.

Really.

An easy way to think of this phenomenon of "regression toward the mean" is that there is a lot more room for variation under the cathedral ceiling in the center of the normal distribution than under the low eaves at the ends.

# Predicted Achievement



Galton, F. (**1886**). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute, 15*, 246-263.

Prediction of a score on one test from a score on another depends largely on the correlation between the two tests. If there were no correlation at all ($r = 0.00$), then the score on the predictor test would be irrelevant.

Assuming you knew nothing else about the examinee, all you could do would be to predict a standard score of 100.

If two tests were perfectly correlated ($r = 1.00$ or $r = -1.00$) you would know the precise score on the predicted test from the score on the predictor.

That prediction raises another issue. If two tests were perfectly correlated ($r = 1.00$ or $r = -1.00$), one test might still give consistently higher scores than the other test, depending on such issues as norming samples.

If the relationship were linear, but one test gave inflated or depressed scores, there would be a 1:1 correspondence between the scores, but 81 might predict 96, 112 might predict 125, and so on.

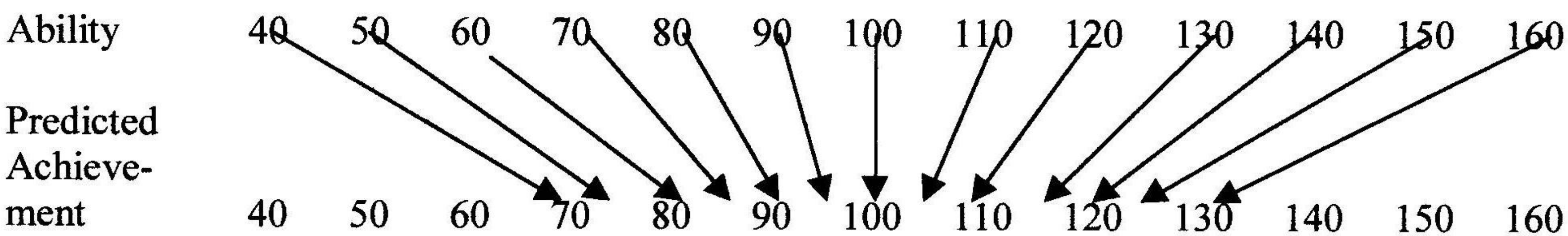
Because all correlations between tests have, so far in history, been higher than 0.00 and less than 1.00, we find that the predicted score falls somewhere between predictor score and the mean (standard score 100).

(In fact all test-retest reliabilities of the same test have been > 0.00 and < 1.00.)

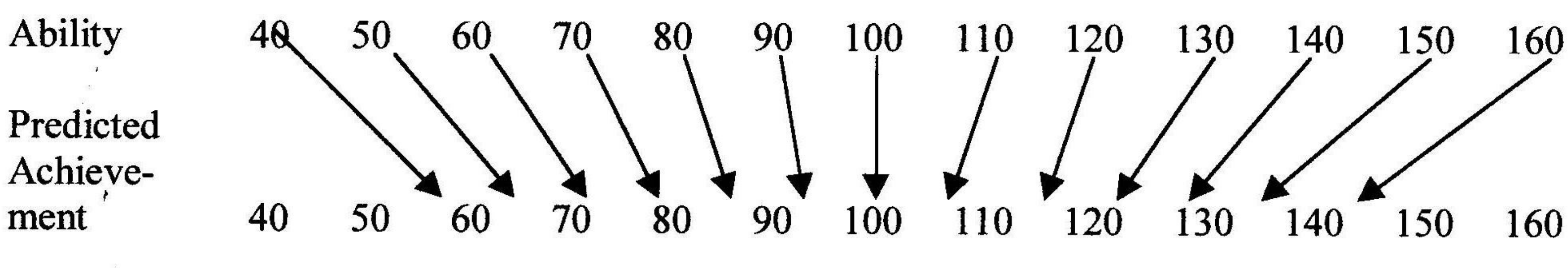
How far the predicted score slips, slides, or regresses toward the mean depends on the correlation between the two tests. You already know that, if the correlation were 0.00, the regression would be total – all the way to the mean.

You also know that, if the correlation were perfect ($\pm$ 1.00), there would be no regression at all.

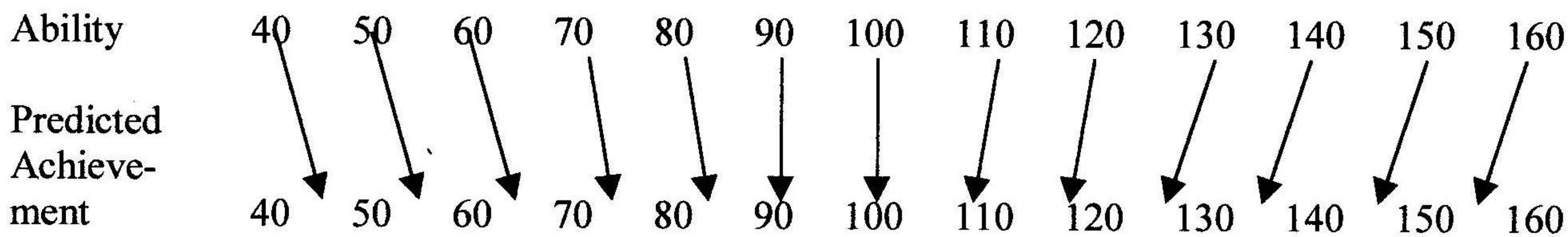
**Achievement Predicted from Ability Scores with a Correlation of .50**



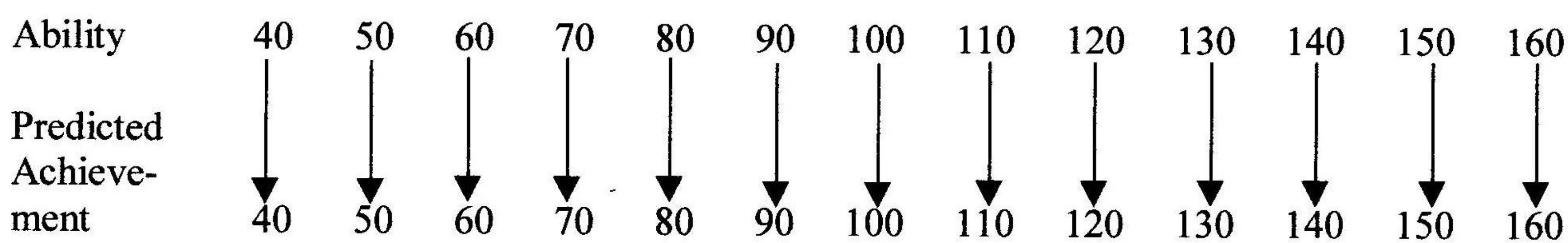

**Achievement Predicted from Ability Scores with a Correlation of .70**



## Achievement Predicted from Ability Scores with a Correlation of .90




## Achievement Predicted from Ability Scores with a Fantasy Correlation of 1.00



It is actually very easy to calculate predicted scores.

You simply convert the predictor score to a z-score (our lingua franca, our Rosetta Stone of statistics), multiply that predictor z score by the correlation, and convert that resulting predicted z score back to whatever statistic (e.g., standard score, scaled score, T score) you want.

You remember that a z-score is the difference between the obtained score (X) and the mean (M) measured in standard deviation (sd) units or

$$z = (X - M)/sd.$$

A standard score of 115 would convert to a z-score as follows:

$$z = (X - M)/sd =$$
$$(115 - 100)/15 =$$
$$(+15)/15 = +1.00$$

A T score of 35 would convert to a z-score as follows:

$$z = (X - M)/sd =$$
$$(35 - 50)/10 =$$
$$(-15)/10 = -1.50$$

You also recall that you convert a z-score to some other statistic by multiplying z by the new standard deviation (e.g., 15 for standard scores or 10 for T scores) and then adding the new mean (e.g., 100 for standard scores or 50 for T scores).

Our z-score of 1.00 would become a scaled score[2] of 13:

new (scaled) score =

$z * sd + M =$

$1.00 * 3 + 10 =$

$3 + 10 = 13$

[2] M = 10, sd = 3, e.g., Wechsler subtest scaled scores.  There are other kinds of scaled scores, but I don't care right now.

Our z-score of –1.50 would become a standard score of 78:

new score =

$z * sd + M =$

$-1.50 * 15 + 100 =$

$-22.5 + 100 =$

$77.5 \approx 78.$

Below is a table illustrating the calculation of predicted scores.  First there are eight predictions with standard scores of 85 and correlations ranging from +.10 to +.90. Then there are eight more with standard scores of 115.  Finally there are examples with different types of scores.

| Predictor Score X | Mean Mx | Standard Deviation SDx | (X-M)/sd z | Correlation r | Predicted z z*r | Mean My | Standard Deviation SDy | Predicted Score Y |
|---|---|---|---|---|---|---|---|---|
| 85 | 100 | 15 | -1.00 | 0.90 | -0.90 | 100 | 15 | 87 |
| 85 | 100 | 15 | -1.00 | 0.80 | -0.80 | 100 | 15 | 88 |
| 85 | 100 | 15 | -1.00 | 0.60 | -0.60 | 100 | 15 | 91 |
| 85 | 100 | 15 | -1.00 | 0.50 | -0.50 | 100 | 15 | 93 |
| 85 | 100 | 15 | -1.00 | 0.40 | -0.40 | 100 | 15 | 94 |
| 85 | 100 | 15 | -1.00 | 0.30 | -0.30 | 100 | 15 | 96 |
| 85 | 100 | 15 | -1.00 | 0.20 | -0.20 | 100 | 15 | 97 |
| 85 | 100 | 15 | -1.00 | 0.10 | -0.10 | 100 | 15 | 99 |
| 115 | 100 | 15 | 1.00 | 0.90 | 0.90 | 100 | 15 | 114 |
| 115 | 100 | 15 | 1.00 | 0.80 | 0.80 | 100 | 15 | 112 |
| 115 | 100 | 15 | 1.00 | 0.60 | 0.60 | 100 | 15 | 109 |
| 115 | 100 | 15 | 1.00 | 0.50 | 0.50 | 100 | 15 | 108 |
| 115 | 100 | 15 | 1.00 | 0.40 | 0.40 | 100 | 15 | 106 |
| 115 | 100 | 15 | 1.00 | 0.30 | 0.30 | 100 | 15 | 105 |
| 115 | 100 | 15 | 1.00 | 0.20 | 0.20 | 100 | 15 | 103 |
| 115 | 100 | 15 | 1.00 | 0.10 | 0.10 | 100 | 15 | 102 |
| Standard Score 68 to T Score with r = .59 | | | | | | | | |
| 68 | 100 | 15 | -2.13 | 0.59 | -1.26 | 50 | 10 | 37 |
| Standard Score 118 to Scaled Score with r = .76 | | | | | | | | |
| 118 | 100 | 15 | 1.20 | 0.76 | 0.91 | 10 | 3 | 13 |
| Scaled Score 11 to T Score with r = .81 | | | | | | | | |
| 11 | 10 | 3 | 0.33 | 0.81 | 0.27 | 50 | 10 | 53 |
| T Score 41 to Scaled Score with r = .88 | | | | | | | | |
| 41 | 50 | 10 | -0.90 | 0.88 | -0.79 | 10 | 3 | 8 |

You may have noticed that the standard deviation seems redundant in the examples in which the beginning and ending scores are the same (e.g., standard scores to standard scores).

You are right: if the predictor and predicted are the same type of score, we can skip the step of dividing by the standard deviation.  You can see that the predicted scores are the same as in the table above.

| Predictor Score X | Mean Mx | Diff from Mean X - M | Correlation r | Diff * Correlation (X-M)*r | Predicted Score Y |
|---|---|---|---|---|---|
| 85 | 100 | -15.00 | 0.90 | -13.50 | 87 |
| 85 | 100 | -15.00 | 0.80 | -12.00 | 88 |
| 85 | 100 | -15.00 | 0.60 | -9.00 | 91 |
| 85 | 100 | -15.00 | 0.50 | -7.50 | 93 |
| 85 | 100 | -15.00 | 0.40 | -6.00 | 94 |
| 85 | 100 | -15.00 | 0.30 | -4.50 | 96 |
| 85 | 100 | -15.00 | 0.20 | -3.00 | 97 |
| 85 | 100 | -15.00 | 0.10 | -1.50 | 99 |
| 115 | 100 | 15.00 | 0.90 | 13.50 | 114 |
| 115 | 100 | 15.00 | 0.80 | 12.00 | 112 |
| 115 | 100 | 15.00 | 0.60 | 9.00 | 109 |
| 115 | 100 | 15.00 | 0.50 | 7.50 | 108 |
| 115 | 100 | 15.00 | 0.40 | 6.00 | 106 |
| 115 | 100 | 15.00 | 0.30 | 4.50 | 105 |
| 115 | 100 | 15.00 | 0.20 | 3.00 | 103 |
| 115 | 100 | 15.00 | 0.10 | 1.50 | 102 |

Prof. Catherine Fiorello observes that this computational shortcut should allow you to calculate predicted achievement without paper and pencil should the issue come up during a team meeting, for example.

That is a pretty cool parlor trick. If you do not know the correlation between an ability test and an achievement test, .60 or .70 would be a reasonable guess for your on-the-spot mental calculation to amaze all the other kids on the team.

When they look up the actual predicted score in one of the test manuals, they will see you were pretty close.  You might be made the witch doctor or the ruler of the IEP tribe.

# To Review:

To predict one score from another, just multiply the predictor z score by the correlation between the tests.

$$z_2 = z_1 * r$$

$$z_2 = z_1 * r$$

FUBAR IQ = 60
correlation (r) with SNAFU
Achievement = 0.70

$$z_2 = z_1 * r$$

FUBAR IQ = 60

FUBAR $z_1$ = (60 – 100)/15

= -40/15 =

= -2.67

$$z_2 = z_1 * r$$

FUBAR $z_1 = -2.67$

SNAFU $z_2 = -2.67 * .70$

$= -1.869$

$$\text{Stand. Score} = z_2 * 15 + 100$$

$$\text{SNAFU } z_2 = -1.869$$

$$\text{SNAFU} = -1.87 * 15 + 100$$
$$= -28 + 100$$
$$= 72$$

FUBAR IQ = 60

r = .70

predicted SNAFU = 72

Predicted Achievement with **r = .50**

| Ability | | | Prediction | |
|---|---|---|---|---|
| score | z | | z | score |
| 40 | -4.00 | | -2.00 | 70 |
| 50 | -3.33 | | -1.67 | 75 |
| 60 | -2.67 | | -1.33 | 80 |
| 70 | -2.00 | | -1.00 | 85 |
| 80 | -1.33 | | -0.67 | 90 |
| 90 | -0.67 | | -0.33 | 95 |
| 100 | 0.00 | | 0.00 | 100 |
| 110 | 0.67 | | 0.33 | 105 |
| 120 | 1.33 | | 0.67 | 110 |
| 130 | 2.00 | | 1.00 | 115 |
| 140 | 2.67 | | 1.33 | 120 |
| 150 | 3.33 | | 1.67 | 125 |
| 160 | 4.00 | | 2.00 | 130 |

Predicted Achievement with **r = .70**

| Ability | | | Prediction | |
|---|---|---|---|---|
| score | z | | z | score |
| 40 | -4.00 | | -2.80 | 58 |
| 50 | -3.33 | | -2.33 | 65 |
| 60 | -2.67 | | -1.87 | 72 |
| 70 | -2.00 | | -1.40 | 79 |
| 80 | -1.33 | | -0.93 | 86 |
| 90 | -0.67 | | -0.47 | 93 |
| 100 | 0.00 | | 0.00 | 100 |
| 110 | 0.67 | | 0.47 | 107 |
| 120 | 1.33 | | 0.93 | 114 |
| 130 | 2.00 | | 1.40 | 121 |
| 140 | 2.67 | | 1.87 | 128 |
| 150 | 3.33 | | 2.33 | 135 |
| 160 | 4.00 | | 2.80 | 142 |

Predicted Achievement with **r = .90**

| Ability | | | Prediction | |
|---|---|---|---|---|
| score | z | | z | score |
| 40 | -4.00 | | -3.60 | 46 |
| 50 | -3.33 | | -3.00 | 55 |
| 60 | -2.67 | | -2.40 | 64 |
| 70 | -2.00 | | -1.80 | 73 |
| 80 | -1.33 | | -1.20 | 82 |
| 90 | -0.67 | | -0.60 | 91 |
| 100 | 0.00 | | 0.00 | 100 |
| 110 | 0.67 | | 0.60 | 109 |
| 120 | 1.33 | | 1.20 | 118 |
| 130 | 2.00 | | 1.80 | 127 |
| 140 | 2.67 | | 2.40 | 136 |
| 150 | 3.33 | | 3.00 | 145 |
| 160 | 4.00 | | 3.60 | 154 |

Note:

Such predictions assume that both tests would give the same score (100) to the average student and both have normal distributions.  If one gives inflated or depressed scores, all bets are off.

**WISC-IV FSIQ 87**

| Subtest or Composite | Predicted Achieve-ment | Actual Score | Difference | Required Difference .01 level | Signifi-cant? | Base Rate |
|---|---|---|---|---|---|---|
| Word Reading | 90 | 104 | 14 | 9.21 | Y | **5%-10%** |
| Numerical Operations | 91 | 83 | -8 | 11.24 | N | |
| Reading Comprehension | 90 | 79 | -11 | 10.04 | Y | 15% |
| Spelling | 91 | 77 | -14 | 11.21 | Y | **10%** |
| Pseudoword Decoding | 92 | 95 | 3 | 7.92 | N | |
| Math Reasoning | 90 | 71 | -19 | 11.53 | Y | **2%** |
| Written Expression | 91 | 79 | -12 | 15.31 | N | |
| Listening Comprehension | 90 | 81 | -9 | 17.70 | N | |
| Oral Expression | 93 | 89 | -4 | 15.44 | N | |
| Reading Composite | 90 | 90 | 0 | 7.66 | N | |
| Math Composite | 90 | 75 | -15 | 9.30 | Y | **5%-10%** |
| Written Language Comp. | 90 | 77 | -13 | 14.37 | N | |
| Oral Language Composite | 90 | 82 | -8 | 11.53 | N | |

# Short-Cut Again

If you stick with one type of score, such as standard scores, you can leave out the standard deviation.

FUBAR IQ 140:  140 – 100 = 40

40 * .70 = 28

Predicted SNAFU = 28 + 100 = 128

Predictions are not, of course, limited to predicting achievement from ability. You might want to predict writing skill from a reading test or adaptive behavior from an IQ test.

If you know the correlation between scores on the Comprehensive Test of Writing Plausible Clueless Answers (CTPCA) and college grade-point averages, you can predict GPA from CTPCA scores.

You should also be able to see why we should not get too excited when an examinee with a very low FSIQ has higher scores for WMI and PSI than for VCI, VSI, and FRI or when an examinee with a very high FSIQ has lower scores for

WMI and PSI than for VCI, VSI, and FRI. WMI and PSI have much lower correlations with the FSIQ (and much lower *g* loadings) than do VCI, VSI, and FRI, so WMI and PSI will show more regression toward the mean than will VCI, VSI, and FRI.

Regression also predicts that a retest score will probably be closer to the mean (e.g., 100) than was the original score.  Statistically, this is true, and it helps account for the fact that scores for children with giftedness and with intellectual disability in the validity sections of test manuals are, respectively, lower and higher than you would expect.  The new test is giving a score closer to the mean that did the cognitive ability test that originally identified the child's giftedness or intellectual disabiity.

However, practice effects are likely to raise retest scores. Motivational factors, such as a desire to beat a previous score (as on the SATs), may raise retest scores. Discouragement about a low previous score or resentment of another evaluation might lower scores.

On cognitive ability tests, the Flynn Effect predicts lower scores on newer than on older tests, all other things being equal.  Changes in norms for achievement tests over time are very complex, but do affect scores on tests normed at different times.

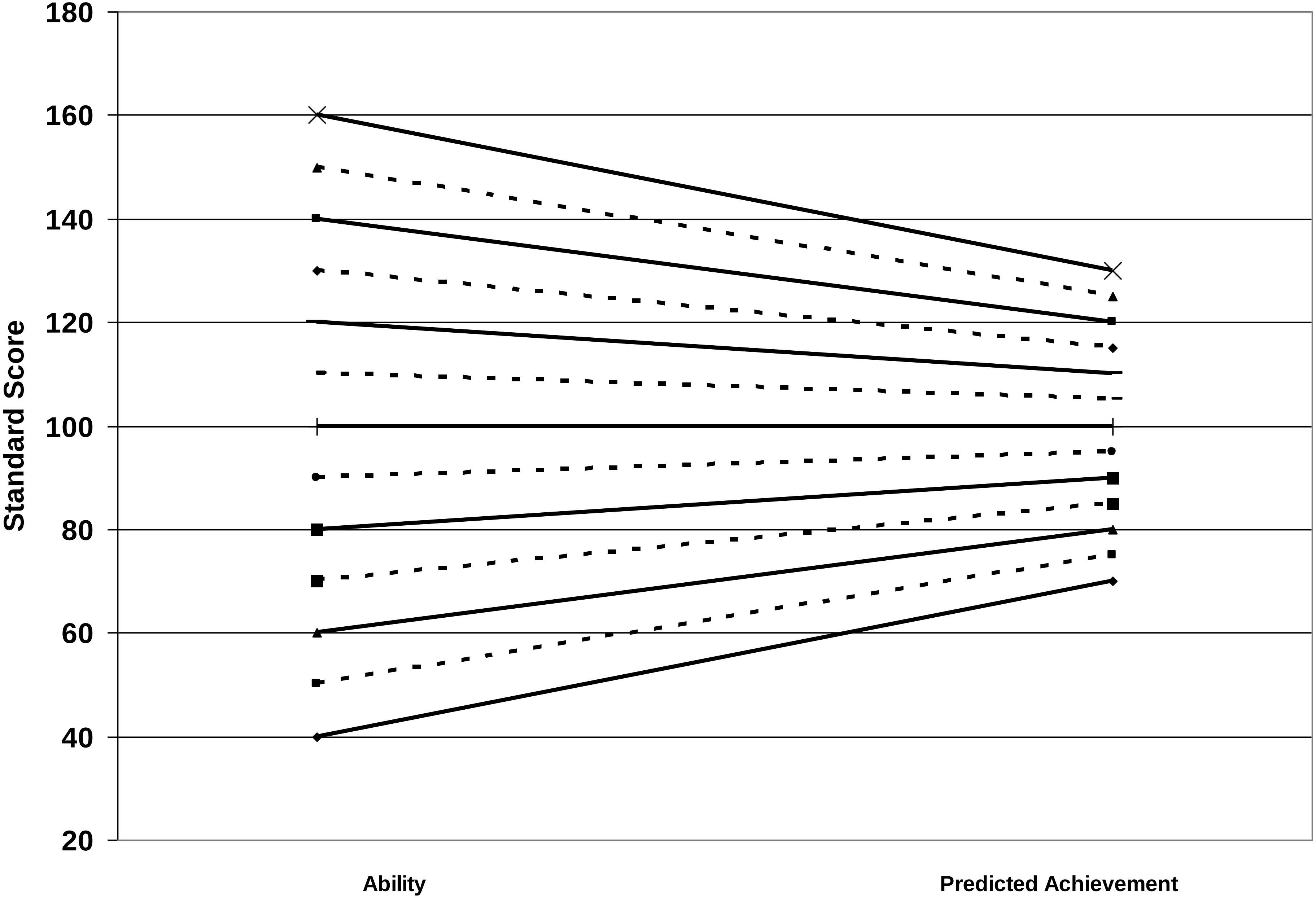
## Noah's KTEA-II Test Scores Compared to Scores Predicted from his WISC-IV Total Score

| | Noah's KTEA-II Score | Score Predicted from WISC-IV | Difference |
|---|---|---|---|
| **Reading** | | | |
| KTEA II A: reading words aloud from a list | 120 | 110 | +10 |
| KTEA II A: speed and accuracy in reading words aloud | 127 | 109 | +18 |
| KTEA II A: speed and accuracy in reading nonsense words aloud | 121 | 108 | +13 |
| KTEA II A: answering comprehension questions about paragraphs | 120 | 109 | +11 |
| **Rapid Automatized Naming and Retrieval Speed** | | | |
| KTEA II A: speed of naming rows of pictures, colors, and letters. | 106 | 105 | + 1 |
| KTEA II A: speed of naming things in specific categories | 89 | 106 | – 17 |
| **Writing** | | | |
| KTEA II A: written spelling of dictated words | 133 | 107 | +26 |
| KTEA II A: writing words, sentences, and an essay | 111 | 107 | + 4 |
| **Math** | | | |
| KTEA II A: math computation with paper and pencil | 103 | 105 | – 2 |
| KTEA II A: math applications ("story" or "word" problems) | 108 | 111 | – 3 |
| | | | |

# r = .50

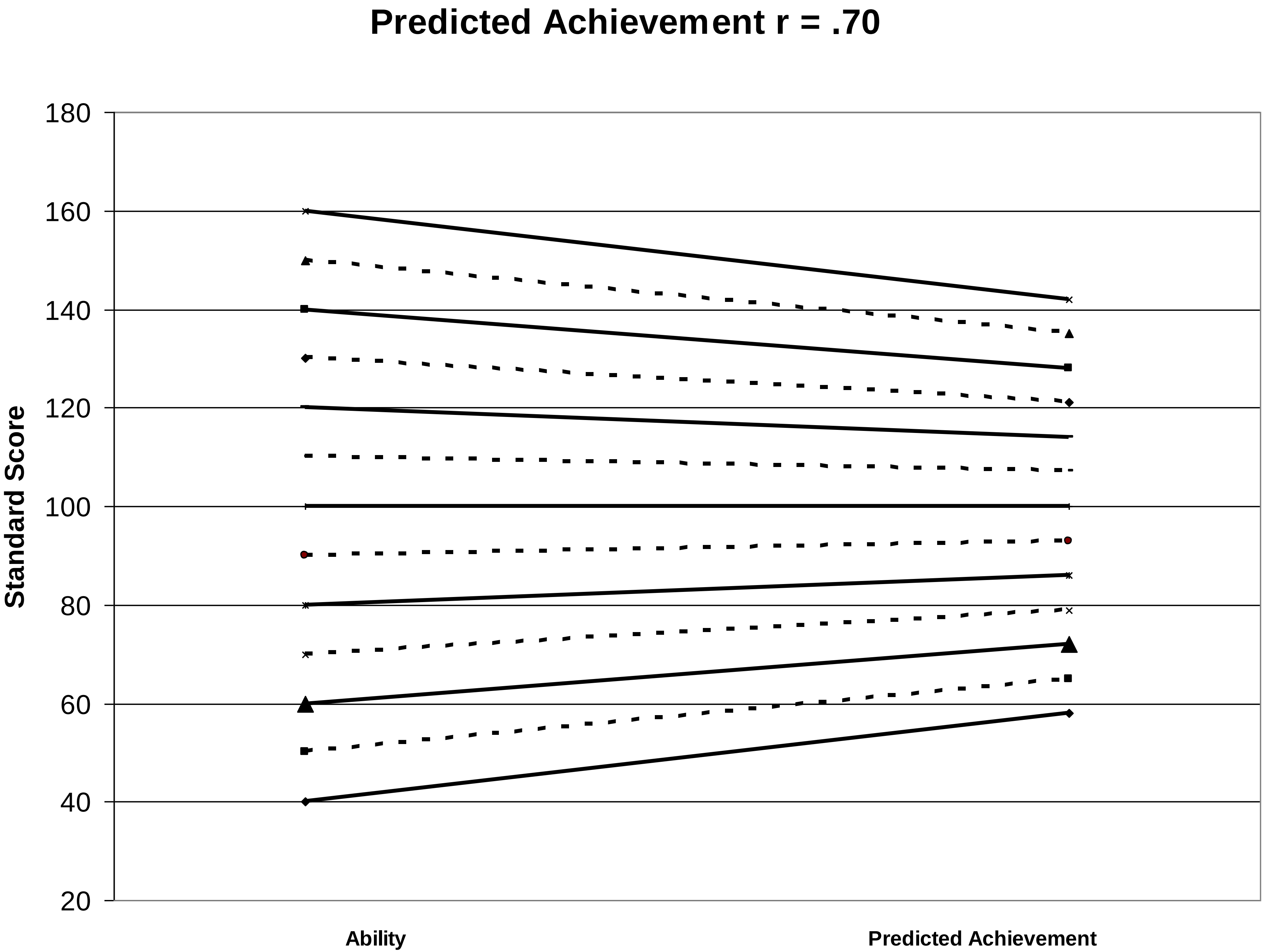
| ABILITY | | PREDICTED | |
|---|---|---|---|
| score | z | z | score |
| 40 | -4.00 | -2.00 | 70 |
| 50 | -3.33 | -1.67 | 75 |
| 60 | -2.67 | -1.33 | 80 |
| 70 | -2.00 | -1.00 | 85 |
| 80 | -1.33 | -0.67 | 90 |
| 90 | -0.67 | -0.33 | 95 |
| 100 | 0.00 | 0.00 | 100 |
| 110 | 0.67 | 0.33 | 105 |
| 120 | 1.33 | 0.67 | 110 |
| 130 | 2.00 | 1.00 | 115 |
| 140 | 2.67 | 1.33 | 120 |
| 150 | 3.33 | 1.67 | 125 |
| 160 | 4.00 | 2.00 | 130 |

# Predicted Achievement r = .50



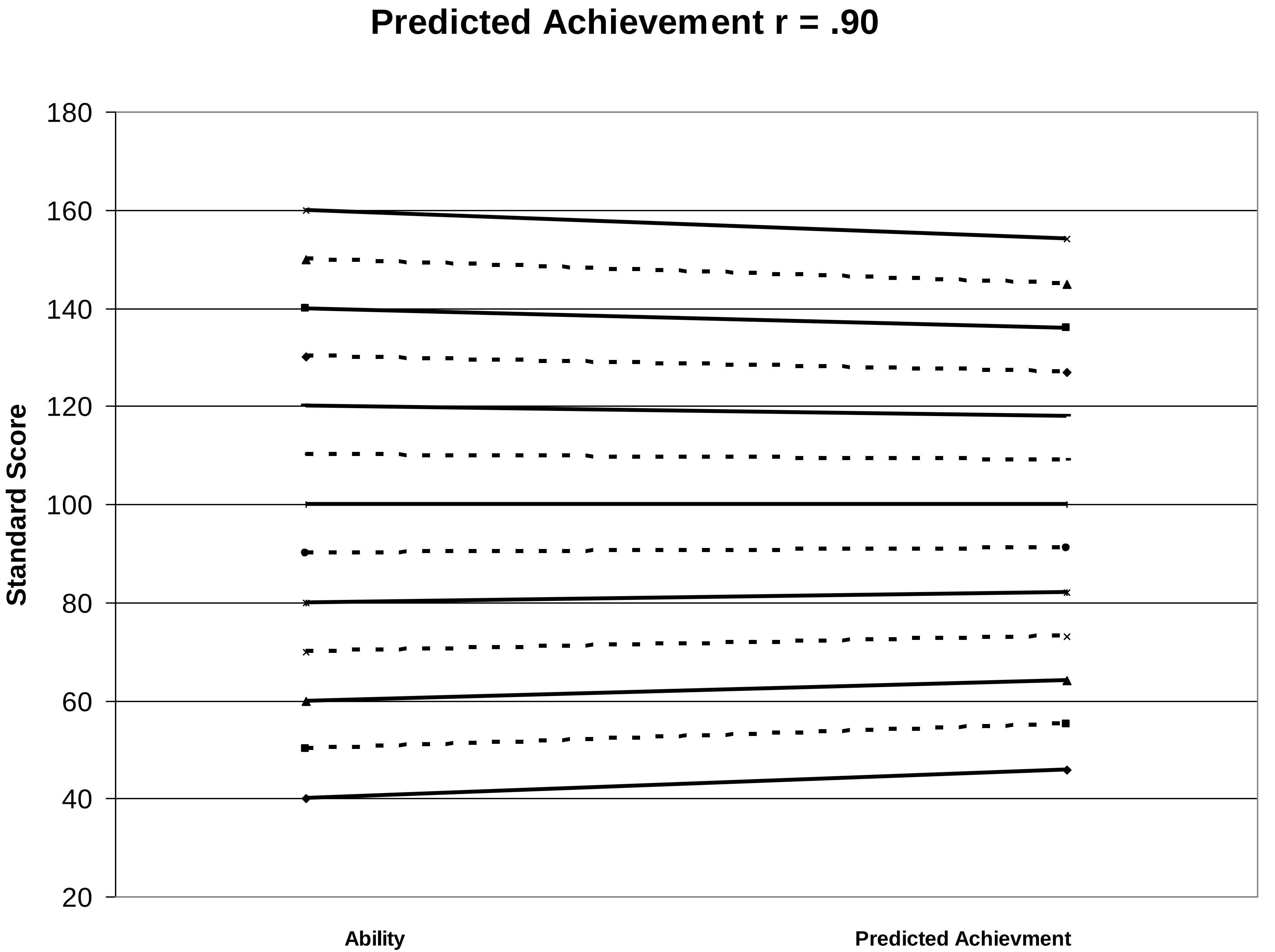
# r = .70

| ABILITY | | PREDICTED | |
|---|---|---|---|
| score | z | z | score |
| 40 | -4.00 | -2.80 | 58 |
| 50 | -3.33 | -2.33 | 65 |
| 60 | -2.67 | -1.87 | 72 |
| 70 | -2.00 | -1.40 | 79 |
| 80 | -1.33 | -0.93 | 86 |
| 90 | -0.67 | -0.47 | 93 |
| 100 | 0.00 | 0.00 | 100 |
| 110 | 0.67 | 0.47 | 107 |
| 120 | 1.33 | 0.93 | 114 |
| 130 | 2.00 | 1.40 | 121 |
| 140 | 2.67 | 1.87 | 128 |
| 150 | 3.33 | 2.33 | 135 |
| 160 | 4.00 | 2.80 | 142 |


Predicted Achievement r = .70
Standard Score
180
160
140
120
100
80
60
40
20
Ability
Predicted Achievement

# r = .90

| ABILITY | | PREDICTED | |
|---|---|---|---|
| score | z | z | score |
| 40 | -4.00 | -3.60 | 46 |
| 50 | -3.33 | -3.00 | 55 |
| 60 | -2.67 | -2.40 | 64 |
| 70 | -2.00 | -1.80 | 73 |
| 80 | -1.33 | -1.20 | 82 |
| 90 | -0.67 | -0.60 | 91 |
| 100 | 0.00 | 0.00 | 100 |
| 110 | 0.67 | 0.60 | 109 |
| 120 | 1.33 | 1.20 | 118 |
| 130 | 2.00 | 1.80 | 127 |
| 140 | 2.67 | 2.40 | 136 |
| 150 | 3.33 | 3.00 | 145 |
| 160 | 4.00 | 3.60 | 154 |


Predicted Achievement r = .90
Standard Score
180
160
140
120
100
80
60
40
20
Ability
Predicted Achievment

# r = 1.00

| ABILITY | | PREDICTED | |
|---|---|---|---|
| score | z | z | score |
| 40 | -4.00 | -4.00 | 40 |
| 50 | -3.33 | -3.33 | 50 |
| 60 | -2.67 | -2.67 | 60 |
| 70 | -2.00 | -2.00 | 70 |
| 80 | -1.33 | -1.33 | 80 |
| 90 | -0.67 | -0.67 | 90 |
| 100 | 0.00 | 0.00 | 100 |
| 110 | 0.67 | 0.67 | 110 |
| 120 | 1.33 | 1.33 | 120 |
| 130 | 2.00 | 2.00 | 130 |
| 140 | 2.67 | 2.67 | 140 |
| 150 | 3.33 | 3.33 | 150 |
| 160 | 4.00 | 4.00 | 160 |

# $r = 1.00$

What you see is what you get (**if** the relationship is linear **and** the tests have the same mean for the same group of people).  When inspecting correlations between tests in the "validity" section of the test manual, be sure to see if the means of the two tests are the same for the same group.

# Caveat!

As Kevin McGrew has warned us in his wonderful Forrest Gump Powerpoint (http://www.iapsych.com/iqach.pdf) (also *Forrest Gump* in your SAIF Handouts and Notes, Assessment of Intellectual Disability folder), half of examinees will score above their predicted achievement, half will score below their predicted achievement, and some will score far above or below the prediction.  Statistical prediction is real and valid, but it is not destiny!

# Further Information

W. Joel Schneider has a fabulous 42-minute discussion of regression to the mean at https://assessingpsyche.wordpress.com/2013/12/16/video-tutorial-misunderstanding-regression-to-the-mean/

Watch it with your sound volume turned up.

# Take-Home Lessons

- When we predict one variable from another, the predicted score will be closer to the mean (more nearly average) than was the predictor score.
- There will be exceptions, sometimes dramatic, but that is the way to bet.
- Therefore, a student with IQ 65 and achievement 65 is **<u>NOT</u>** performing to expectations.  That student is achieving far below expectations!

# Take-Home Lessons

- Identifying specific learning disabilities by using statistical formulae involving "ability" and "achievement" tests is a stupid practice.
- If you are forced to do it, you must use predicted achievement scores (now found in most test manuals).
- Simple differences (e.g., 15 or 22.5 points) are indefensible.  They will over-identify students with high IQs and under-identify students with below average IQs.