Psychology 405: Psychometric Theory

Validity

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

May, 2018
Outline

- Preliminaries
- Predictions and Decisions
- The VA study
Observed Variables

\[ X \]

\[ X_1 \]
\[ X_2 \]
\[ X_3 \]
\[ X_4 \]
\[ X_5 \]
\[ X_6 \]

\[ Y \]

\[ Y_1 \]
\[ Y_2 \]
\[ Y_3 \]
\[ Y_4 \]
\[ Y_5 \]
\[ Y_6 \]
Latent Variables

$\xi$ $\eta$

$\xi_1$ $\eta_1$

$\xi_2$ $\eta_2$
Theory: A regression model of latent variables

\[ \xi \quad \eta \]

\[ \xi_1 \quad \eta_1 \]

\[ \xi_2 \quad \eta_2 \]

\[ \zeta_1 \quad \zeta_2 \]
A measurement model for \( X \) – Correlated factors

\[
\delta \xrightarrow{} X \xrightarrow{} \xi
\]

\[\delta_1 \xrightarrow{} X_1\]
\[\delta_2 \xrightarrow{} X_2 \xrightarrow{} \xi_1\]
\[\delta_3 \xrightarrow{} X_3 \xrightarrow{} \xi_2\]
\[\delta_4 \xrightarrow{} X_4 \xrightarrow{} \xi_1\]
\[\delta_5 \xrightarrow{} X_5 \xrightarrow{} \xi_2\]
\[\delta_6 \xrightarrow{} X_6\]
A measurement model for $Y$ - uncorrelated factors

\[ \eta \rightarrow Y \rightarrow \epsilon \]

\[ \eta_1 \rightarrow Y_1 \rightarrow \epsilon_1 \]
\[ \eta_1 \rightarrow Y_2 \rightarrow \epsilon_2 \]
\[ \eta_1 \rightarrow Y_3 \rightarrow \epsilon_3 \]
\[ \eta_2 \rightarrow Y_4 \rightarrow \epsilon_4 \]
\[ \eta_2 \rightarrow Y_5 \rightarrow \epsilon_5 \]
\[ \eta_2 \rightarrow Y_6 \rightarrow \epsilon_6 \]
A complete structural model

\[ \delta \quad X \quad \xi \quad \eta \quad Y \quad \epsilon \]

\[ \delta_1 \rightarrow X_1 \rightarrow \xi_1 \rightarrow \eta_1 \rightarrow Y_1 \rightarrow \epsilon_1 \]
\[ \delta_2 \rightarrow X_2 \rightarrow \xi_1 \rightarrow \eta_1 \rightarrow Y_2 \rightarrow \epsilon_2 \]
\[ \delta_3 \rightarrow X_3 \rightarrow \xi_1 \rightarrow \eta_1 \rightarrow Y_3 \rightarrow \epsilon_3 \]
\[ \delta_4 \rightarrow X_4 \rightarrow \xi_1 \rightarrow \eta_1 \rightarrow Y_4 \rightarrow \epsilon_4 \]
\[ \delta_5 \rightarrow X_5 \rightarrow \xi_1 \rightarrow \eta_1 \rightarrow Y_5 \rightarrow \epsilon_5 \]
\[ \delta_6 \rightarrow X_6 \rightarrow \xi_1 \rightarrow \eta_1 \rightarrow Y_6 \rightarrow \epsilon_6 \]
Types of Validity

- **Face/Faith**
  - $X_1$ to $Y_1$
  - $X_2$ to $Y_2$

- **Concurrent**
  - $X_2$ to $Y_2$
  - $X_3$ to $Y_3$

- **Predictive**
  - $X_3$ to $Y_3$
  - $X_4$ to $Y_4$

- **Convergent**
  - $X_4$ to $Y_4$
  - $X_5$ to $Y_5$

- **Discriminant**
  - $X_5$ to $Y_5$
  - $X_6$ to $Y_6$

- **Construct**
  - $X_6$ to $Y_6$
  - $X_7$ to $Y_7$
  - $X_8$ to $Y_8$
Face Validity

$X_1$

*Face/Faith*

Representative Content

Seeming relevance
Concurrent Validity

Does a measure correlate with the criterion?

Need to define the criterion.

Assumes that what correlates now will have predictive value.
Predictive Validity

$X_3 \xrightarrow{\text{Predictive}} Y_3$

Does a measure correlate with the criterion?

Need to define the criterion.

Allow time to pass
Prediction

1. Continuous predictor, continuous criterion
   - Regression, multiple regression, correlation
   - Slope of regression implies how much change for unit change in predictor

2. Continuous predictor, dichotomous criterion
   - point bi-serial correlation

3. Dichotomous predictor, dichotomous outcome
   - Phi
   - The Taylor-Russell tables (Taylor & Russell, 1939) and the problem of Selection Ratios and Base Rates

\[
\phi = \frac{VP - BR \times SR}{\sqrt{(BR)(1 - BR)(SR)(1 - SR)}} \tag{1}
\]

- Therefore, the number of valid positives is

\[
VP = BR \times SR + \phi \sqrt{(BR)(1 - BR)(SR)(1 - SR)} \tag{2}
\]
Tetrachoric and phi as function of cut points

\[ \rho = 0.4 \]
\[ \phi = 0.26 \]
A decision theoretic approach

Valid Positives as function of False Positives

Decision Theory
Tetrachoric and phi as function of cut points .5,0

\[ \rho = 0.4 \]
\[ \phi = 0.25 \]
A decision theoretic approach with low beta

Valid Positives as function of False Positives

Decision Theory
Tetrachoric and phi as function of cut points 1,0

\[ \rho = 0.4 \]
\[ \phi = 0.21 \]
A decision theoretic approach with high beta

Valid Positives as function of False Positives

Decision Theory
Tetrachoric and phi as function of cut points 2,0

\[ Y \rho = 0.4 \]
\[ \phi = 0.11 \]
A decision theoretic approach with high beta

Valid Positives as function of False Positives

Decision Theory

Probability of observation
A decision theoretic analysis with 4 different cut points

Valid Positives as function of False Positives

Decision Theory

Probability of observation

Valid Positives as function of False Positives

Decision Theory

Valid Positives as function of False Positives

Decision Theory

Valid Positives as function of False Positives

Decision Theory

Valid Positives as function of False Positives

Decision Theory

Valid Positives as function of False Positives

Decision Theory

Valid Positives as function of False Positives

Decision Theory

Valid Positives as function of False Positives

Decision Theory

Valid Positives as function of False Positives

Decision Theory

Valid Positives as function of False Positives

Decision Theory

Valid Positives as function of False Positives

Decision Theory
Applying decision theory to a prediction problem: the case of predicting future psychiatric diagnoses from military inductees. (Data from Danielson & Clark (1954) as discussed by Wiggins (1973).

<table>
<thead>
<tr>
<th></th>
<th>Predicted Positive</th>
<th>Predicted Negative</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive</td>
<td>49</td>
<td>40</td>
<td>99</td>
</tr>
<tr>
<td>True Negative</td>
<td>79</td>
<td>336</td>
<td>406</td>
</tr>
<tr>
<td>Column Totals</td>
<td>118</td>
<td>376</td>
<td>505</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Fraction of Total</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive</td>
<td>.097</td>
<td>.079</td>
<td>.196</td>
</tr>
<tr>
<td>True Negative</td>
<td>.157</td>
<td>.667</td>
<td>.804</td>
</tr>
<tr>
<td>Column Totals</td>
<td>.234</td>
<td>.746</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Accuracy = .097 + .667 = .76
Sensitivity = .097 / (.097 + .079) = .55
Specificity = .667 / (.667 + .157) = .81
Phi = \[
\frac{.097 - .196 \times .234}{\sqrt{.196 \times .804 \times .234 \times .747}} = .32
\]
The Danielson and Clark data set as a decision problem

Valid Positives as function of False Positives

Decision Theory

Probability of observation

24 / 48
Classics in Prediction and selection

1. Gideon’s selection of soldiers
2. OSS and Army Air Corps selection studies
3. ? (1950) selection of psychology students (?)
4. Astronaut selection
5. Peace Corps selection
Gideon’s assessment
The assessment of pilots – how to show a .45 correlation makes a difference
Predicting clinical psychologists – Kelly and Fiske

1. Multiple predictors of graduate school performance: Kelly and Fiske (1950), Multiple predictors
2. Ability, Interests, temperament (each with $r \approx 0.2 - 0.25$) have multiple R of 0.4–0.5
3. Are they able, interested and stable?
VA study: overview

- Researchers
  - nearly 40 cooperating clinical training programs
  - \( \approx 75 \) psychologists on research staff
- Participants
  - 3/4 of those entering graduate training in 1946, 1947, 1948
  - \( N = 160, 128, 545 \) (selected down to 98)
- Measures
  - Objective tests
  - Clinical assessments
Objective instruments

- **Ability**
  - Millers Analogy Test
  - Thurstone Tests of Primary Mental Abilities

- **Temperament and Character**
  - Minnesota Multiphasic Personality Inventory
  - Guildord Martin Battery of Personality Inventories

- **Interests, Values**
  - Allport-Vernon Scale of Values
  - Strong Vocational Interest Blank
  - Kuder Preference Record
Assessment ratings

• Seven days of tests, interviews and “other” procedures
  • Three raters spent a week studying 4 trainees
  • Staff time devoted to each candidate was at least 7 man-days

• Ratings based on interviews, projective tests, role playing
  • Ratings on:
    • 22 descriptive variables (e.g., cooperativeness, talkativeness)
    • 10 evaluative variables (e.g., social adjustment, emotional expression)
    • 11 predictive variables (e.g. academic, diagnostician, overall suitability)
Criterion variables after 2 years

- Training status (Failure, still in Training, Ph.D. obtained)
- 2nd year evaluations
  - Skill in clinical diagnosis
  - Skill in individual psychotherapy
  - Skill in Research
  - Preference for hiring
- Generally high correlations among all the criteria
High correlations among the criteria

Intercorrelations among selected criterion evaluations

N = 130 P-3 trainees evaluated in the spring of 1949, for whom all evaluation measures were available.

<table>
<thead>
<tr>
<th></th>
<th>Clinical Diagnosis</th>
<th>Individual Therapy</th>
<th>Research</th>
<th>Preference for Hiring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univ.¹</td>
<td>Univ.</td>
<td>Univ.</td>
<td>Univ.</td>
</tr>
<tr>
<td>Clinical Diagnosis:</td>
<td>72</td>
<td>81</td>
<td>55</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>79</td>
<td>60</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Individual Therapy:</td>
<td>62</td>
<td>73</td>
<td>48</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>63</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Research:</td>
<td>11</td>
<td>31</td>
<td>65</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>31</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Preference for Hiring:</td>
<td>44</td>
<td>70</td>
<td>31</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>40</td>
<td>56</td>
<td></td>
</tr>
</tbody>
</table>

¹ University staff evaluations.
² Installation staff evaluations.
The more they know about you, the more they will judge you

<table>
<thead>
<tr>
<th>Assessor</th>
<th>Information on Which Predictions Were Based</th>
<th>Criterion Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Clinical Diagnosis</td>
</tr>
<tr>
<td>Projectivist</td>
<td></td>
<td>02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-01</td>
</tr>
<tr>
<td>Proj. Integration</td>
<td></td>
<td>-05</td>
</tr>
<tr>
<td>Initial Interviewer</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Initial Interviewer</td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>Intensive Interviewer</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Intensive Interviewer</td>
<td></td>
<td>36**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24*</td>
</tr>
<tr>
<td>Pre-Conference</td>
<td></td>
<td>38*</td>
</tr>
<tr>
<td>Situationists (Pooled Rating)</td>
<td></td>
<td>30**</td>
</tr>
<tr>
<td>Prelim. Pooled</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>Final Pooled</td>
<td></td>
<td>23*</td>
</tr>
</tbody>
</table>
### Objectives are just as good

#### B. Objective Test Scores

<table>
<thead>
<tr>
<th>Test Type</th>
<th>24*</th>
<th>15</th>
<th>06</th>
<th>05</th>
<th>24*</th>
<th>23*</th>
<th>23*</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Miller Analogies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Strong Test</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychologist—1938</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychologist—1948 (Kriedt)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychologist, Clinical, 1948 (Kriedt)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychologist, VA Clinical (This Project)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allport-Vernon Theoretical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guilford-Martin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C—Lack of Cycloid Disposition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N—Lack of Nervous Tenseness and Irritability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the finding that the interview did not add to, but actually tended to decrease, the validity of clinical judgments made in the 1947 assessment program was confirmed by submitting the paper and-pencil materials on these same candidates to a later assessment staff which made predictions without any face-to-face contact with the assessee. Under these conditions, the new staff made predictions with slightly higher validities than those made by the staff in 1947, who had the additional data from the interview, situation tests, etc.
Interests matter

The VA Clinical Psychologist key, developed by this project on the basis of the responses of full time VA psychologists, regularly yields relatively high correlations with all criterion evaluations, and compares favorably with the best predictions based on assessment ratings. Other psychologist keys, including the original (1938) general psychologist key and two developed by Kriedt (2), do fairly well. Not shown in the table is a correlation of .61 (N = 44) between scores based on the psychologist key (1938) and the scores made on the objective test of Knowledge of Clinical Psychology three years later. Thus, scores from a single objective test obtainable by mail, at little cost, predicted each of several criteria as well as any of the clinical judgments made in the entire assessment program.
Motivation

Our findings suggest that, in selection for professional training, more attention might well be given to the role of motivation. Perhaps at the level of graduate training, we need establish only a minimal cutting score on tests of intellectual aptitudes; beyond that point, the strength of motivation and the absence of conflicting drives may be the determining factors in success in professional training, and even in the conduct of professional duties.
Faith validity of interviews

Many who have seen our results have been disturbed by the findings regarding the validity for this selection problem of specific techniques which are felt by many professional psychologists to have a high degree of face-validity (or is it faith validity?). Thus, it was the firm conviction of the staff of the OSS assessment program that the global evaluation of a person permits much more accurate predictions of his future performance than can possibly be achieved by a more segmental approach. Unfortunately, the OSS data did not provide a conclusive answer to this question. Our own findings to date serve to raise doubts concerning the validity of this general proposition.
We must evaluate our judgments

Evidence such as that accumulating in this project serves to remind us of the fallibility of the human being both as a measuring device and as an integrator of data. In laboratories, in factories, and in accounting offices, it has been found necessary to supplement his sensory and perceptual capacities with an elaborate array of measuring instruments and computing devices. Pending the gradual development of better measures of psychological variables and comparable aids for combining them, we must continue to rely heavily on human judgment. In so doing, however, we must be continually aware of the magnitude of the errors of such judgments. These errors can be minimized by placing greatest reliance on measures of demonstrated reliability and validity.
We are, in fact, rather encouraged at the probability of being able to predict such criteria with a multiple R of around .50 on the basis of an inexpensive test battery which may be administered without requiring the applicant to present himself at the university of his choice.
More recent prediction studies

1. Terman & Oden (1947, 1959); Oden (1968)
2. Kuncel, Campbell & Ones (1998); Kuncel, Hezlett & Ones (2001); Kuncel & Hezlett (2007) and graduate school prediction
### Kuncel et al. meta analysis predicting graduate school performance

#### Table 2

*META-ANALYSIS OF GRE AND UGPA VALIDITIES: TOTAL SAMPLE*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>N</th>
<th>k</th>
<th>r_{obs}</th>
<th>SD_{obs}</th>
<th>SD_{res}</th>
<th>( \rho )</th>
<th>SD_{\rho}</th>
<th>90% credibility interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GGPA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>14,156</td>
<td>103</td>
<td>.23</td>
<td>.14</td>
<td>.10</td>
<td>.34 .15</td>
<td>.09 to .59</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>14,425</td>
<td>103</td>
<td>.21</td>
<td>.11</td>
<td>.06</td>
<td>.32 .08</td>
<td>.19 to .45</td>
<td></td>
</tr>
<tr>
<td>Analytical</td>
<td>1,928</td>
<td>20</td>
<td>.24</td>
<td>.12</td>
<td>.04</td>
<td>.36 .06</td>
<td>.26 to .46</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>2,413</td>
<td>22</td>
<td>.31</td>
<td>.12</td>
<td>.05</td>
<td>.41 .07</td>
<td>.30 to .52</td>
<td></td>
</tr>
<tr>
<td>UGPA^*</td>
<td>9,748</td>
<td>58</td>
<td>.28</td>
<td>.13</td>
<td>.10</td>
<td>.30 .11</td>
<td>.12 to .48</td>
<td></td>
</tr>
<tr>
<td><strong>1st-year GGPA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>45,615</td>
<td>1,231</td>
<td>.24</td>
<td>.19</td>
<td>.09</td>
<td>.34 .12</td>
<td>.14 to .54</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>45,618</td>
<td>1,231</td>
<td>.24</td>
<td>.19</td>
<td>.08</td>
<td>.38 .12</td>
<td>.18 to .58</td>
<td></td>
</tr>
<tr>
<td>Analytical</td>
<td>36,325</td>
<td>1,080</td>
<td>.24</td>
<td>.19</td>
<td>.06</td>
<td>.36 .09</td>
<td>.21 to .51</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>10,225</td>
<td>98</td>
<td>.34</td>
<td>.11</td>
<td>.03</td>
<td>.45 .04</td>
<td>.38 to .52</td>
<td></td>
</tr>
<tr>
<td>UGPA^*</td>
<td>42,193</td>
<td>1,178</td>
<td>.30</td>
<td>.18</td>
<td>.10</td>
<td>.33 .10</td>
<td>.17 to .49</td>
<td></td>
</tr>
<tr>
<td><strong>Comprehensive exam scores^b</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal^c</td>
<td>1,198</td>
<td>11</td>
<td>.34</td>
<td>.16</td>
<td>.12</td>
<td>.44 .15</td>
<td>.19 to .69</td>
<td></td>
</tr>
<tr>
<td>Quantitative^c</td>
<td>1,194</td>
<td>11</td>
<td>.19</td>
<td>.11</td>
<td>.04</td>
<td>.26 .06</td>
<td>.16 to .36</td>
<td></td>
</tr>
<tr>
<td>Subject^d</td>
<td>534</td>
<td>4</td>
<td>.43</td>
<td>.07</td>
<td>.00</td>
<td>.51 .00</td>
<td>.51 to .51</td>
<td></td>
</tr>
<tr>
<td>UGPA^e</td>
<td>592</td>
<td>6</td>
<td>.12</td>
<td>.05</td>
<td>.00</td>
<td>.12 .00</td>
<td>.12 to .12</td>
<td></td>
</tr>
<tr>
<td><strong>Faculty ratings</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>4,766</td>
<td>35</td>
<td>.23</td>
<td>.12</td>
<td>.08</td>
<td>.42 .14</td>
<td>.19 to .65</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>5,112</td>
<td>34</td>
<td>.25</td>
<td>.10</td>
<td>.02</td>
<td>.47 .04</td>
<td>.40 to .54</td>
<td></td>
</tr>
<tr>
<td>Analytical</td>
<td>1,982</td>
<td>9</td>
<td>.23</td>
<td>.05</td>
<td>.00</td>
<td>.35 .00</td>
<td>.35 to .35</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>879</td>
<td>12</td>
<td>.30</td>
<td>.16</td>
<td>.11</td>
<td>.50 .18</td>
<td>.20 to .80</td>
<td></td>
</tr>
<tr>
<td>UGPA^e</td>
<td>3,695</td>
<td>22</td>
<td>.25</td>
<td>.12</td>
<td>.10</td>
<td>.35 .14</td>
<td>.12 to .58</td>
<td></td>
</tr>
<tr>
<td><strong>Degree attainment^a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>6,304</td>
<td>32</td>
<td>.14</td>
<td>.14</td>
<td>.12</td>
<td>.18 .16</td>
<td>-.08 to .44</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>6,304</td>
<td>32</td>
<td>.14</td>
<td>.17</td>
<td>.15</td>
<td>.20 .20</td>
<td>-.13 to .53</td>
<td></td>
</tr>
<tr>
<td>Analytical</td>
<td>1,233</td>
<td>16</td>
<td>.08</td>
<td>.25</td>
<td>.22</td>
<td>.11 .33</td>
<td>-.38 to .60</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>2,575</td>
<td>11</td>
<td>.32</td>
<td>.16</td>
<td>.14</td>
<td>.39 .17</td>
<td>.11 to .67</td>
<td></td>
</tr>
<tr>
<td>UGPA^e</td>
<td>6,315</td>
<td>33</td>
<td>.12</td>
<td>.17</td>
<td>.16</td>
<td>.12 .16</td>
<td>-.14 to .38</td>
<td></td>
</tr>
</tbody>
</table>

#### Time to complete^b,c

<table>
<thead>
<tr>
<th>Predictor</th>
<th>N</th>
<th>k</th>
<th>r_{obs}</th>
<th>SD_{obs}</th>
<th>SD_{res}</th>
<th>( \rho )</th>
<th>SD_{\rho}</th>
<th>90% credibility interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>6,304</td>
<td>32</td>
<td>.14</td>
<td>.14</td>
<td>.12</td>
<td>.18 .16</td>
<td>-.08 to .44</td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>6,304</td>
<td>32</td>
<td>.14</td>
<td>.17</td>
<td>.15</td>
<td>.20 .20</td>
<td>-.13 to .53</td>
<td></td>
</tr>
<tr>
<td>Analytical</td>
<td>1,233</td>
<td>16</td>
<td>.08</td>
<td>.25</td>
<td>.22</td>
<td>.11 .33</td>
<td>-.38 to .60</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>2,575</td>
<td>11</td>
<td>.32</td>
<td>.16</td>
<td>.14</td>
<td>.39 .17</td>
<td>.11 to .67</td>
<td></td>
</tr>
<tr>
<td>UGPA^e</td>
<td>6,315</td>
<td>33</td>
<td>.12</td>
<td>.17</td>
<td>.16</td>
<td>.12 .16</td>
<td>-.14 to .38</td>
<td></td>
</tr>
</tbody>
</table>
Kuncel et al. meta analysis predicting graduate school performance

### Table 9

**GRE and UGPA Unit-Weighted Composite Predicting GGPA and Faculty Ratings**

<table>
<thead>
<tr>
<th>Predictor set</th>
<th>Predictive validity of unit-weighted composite</th>
<th>Predictive validity of composite plus UGPA (unit weighted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>.41</td>
<td>.48</td>
</tr>
<tr>
<td>Quantitative</td>
<td>.42</td>
<td>.50</td>
</tr>
<tr>
<td>Analytical</td>
<td>.38</td>
<td>.46</td>
</tr>
<tr>
<td>Subject</td>
<td>.49</td>
<td>.54</td>
</tr>
<tr>
<td>Verbal + Quantitative</td>
<td>.46</td>
<td>.53</td>
</tr>
<tr>
<td>Verbal + Quantitative + Analytical</td>
<td>.45</td>
<td>.50</td>
</tr>
<tr>
<td>Verbal + Quantitative + Subject</td>
<td>.52</td>
<td>.56</td>
</tr>
<tr>
<td>Verbal + Quantitative + Analytical + Subject</td>
<td>.50</td>
<td>.54</td>
</tr>
</tbody>
</table>

*Note.* GRE = Graduate Record Examinations; UGPA = undergraduate grade point average; GGPA = graduate grade point average.
**Benbow and Lubinski: Beyond the threshold**

*Beyond the Threshold Hypothesis*

---

**Outcome**
- Any Doctorate (PhD, MD, JD): OR = 2.7*
- Any Peer-reviewed Publication: OR = 4.5*
- STEM Publications (≥1): OR = 5.9*
- STEM Doctorates: OR = 18.2*
- Patents (≥1): OR = 6.1*
- Income in 95th Percentile: OR = 3.3*
- STEM Tenure (Top 50): OR = 7.7*

---

**Age 13 SAT Math Score**

**Proportion of Quartile With Outcome**

Q4, Q3, Q2, Q1
Figure 3. Scattergram of age-corrected Moray House Test (MHT) scores at age 11 and age 80 for participants in the Lothian Birth Cohort 1921 of the Scottish Mental Survey 1932.
Deary: the Scottish sample and mortality
Figure 6. Some possible influences and pathways linking mental ability in childhood and survival. From *Brain and Longevity: Perspectives in Longevity* (p. 162, Figure 3), by C. Finch, J.-M. Robine, & Y. Christen (Eds.), 2003, Berlin: Springer. Copyright 2003 by Springer. Adapted with permission.


