Noncognitive skills assessment can be improved with innovative new measures

Patrick Kyllonen
Educational Testing Service
Princeton, New Jersey, USA

High Level Policy Forum
Skills for Social Progress
III. Researchers’ Forum: “Measuring Skills that Matter”
Sao Paolo, Brazil
March 25, 2014
Why alternative measures for noncognitive assessment?

• Self-rating scales are very useful as is—most of what we know about noncognitive skills is based on such scales!

• But there are problems with these scales
  – Socially desirable responding (wanting to look good)
  – Reference group bias (who you compare to)
  – Response style bias (e.g., extreme responses, modesty)
  – Cross-cultural comparability (to compare countries x and y)
  – For others’ ratings: Lack of differentiation (halo, horn)
There are methods to address these problems

• Situational judgment tests
• Behaviorally anchored rating scales
• Performance measures
• Ratings by others (teachers, parents)
• Forced-choice assessments
• Anchoring vignettes

But I’ll focus on these two

I’ll say something about this
Others’ Ratings (teachers, parents, friends)

- **Psychologists’ ratings** of 18-year-olds’ noncognitive skills were comparable to or more powerful than IQ in predicting earnings, employment, and chronic unemployment 20 years later (Lindqvist & Vestman, 2011)

- **Teachers’ ratings** of 8th graders’ misbehavior (5-item checklist) were comparable to or better than achievement tests in predicting educational attainment and earnings 20 years later (Segal, 2012)

- **Others ratings** add to & are better than self-ratings in predicting academic achievement & job performance (Connelly & Ones, 2010; Oh, Wang, Mount, 2011)

Even **casual familiarity** allows for accurate ratings, for many dimensions.
FORCED-CHOICE ASSESSMENTS
## Single Statements Rating Scale

Please indicate your answer to each item by clicking on the appropriate circle.

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I keep my promises</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>2. I am generally pretty forgiving</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

## Forced-Choice

For each pair of statements please click on the one that is most like you.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I keep my promises</td>
<td>○</td>
</tr>
<tr>
<td>2. I am generally pretty forgiving</td>
<td>○</td>
</tr>
</tbody>
</table>

Drasgow, Stark, Chernyshenko, Nye, Hulin, & White (2012).
Forced Choice vs. Single Statements

• Forced-choice shows higher validities vs. single statements PISA 2012; Brown & Bartram (2009); Bartram (2013)
  – For example, correlation between conscientiousness and school and job performance: Salgado & Táuriz (2012)
    • Forced-choice: \( r = .40 \)
    • Single Statement Ratings: \( r = .16 \)
  – New approaches use item-response theory scoring of forced-choice data (Stark et al, 2005; Brown & Maydeu-Olivares, 2013)

• Forced-choice provides better cross-cultural comparability vs. single statements (Bartram, 2013)
  • ...next page...
Cross-cultural comparability

<table>
<thead>
<tr>
<th>Country-level correlations (n = 19) between</th>
<th>UN Human Development Index (education, life expectancy, GDP)</th>
<th>Global competitive index (WEF), requirements, efficiency, innovation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreeableness</td>
<td>Single Statement</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>Forced Choice</td>
<td>.57</td>
</tr>
<tr>
<td>Emotional stability</td>
<td>Single Statement</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Forced choice</td>
<td>.27</td>
</tr>
<tr>
<td>Extraversion</td>
<td>Single Statement</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>Forced choice</td>
<td>.76</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>Single Statement</td>
<td>-.46</td>
</tr>
<tr>
<td></td>
<td>Forced Choice</td>
<td>-.08</td>
</tr>
</tbody>
</table>

(Bartram, 2013)

Copyright © 2014 Educational Testing Service. All rights reserved.
ANCHORING VIGNETTES
Cross-Cultural Validity

• Attitude-achievement “paradox”
  – Positive average within country correlations
    • “Better attitudes are associated with higher achievement”
  – Negative country-level correlations
    • “Countries with high average attitude scores are ones with lower average achievement”
    • “Countries with low average attitudes are ones with high achievement”
Correlations with Mathematics Achievement Scores

Within-country Alignment of within-country and country-level correlations

Correlations with Mathematics Achievement Scores

Country-level correlation (60 countries)

Average correlation within country (N = 1,000+ students)
Anchoring Vignettes

• PISA 2012: anchoring vignettes (and forced choice) “solved” this problem

• Anchoring vignettes are a method for rescaling Likert scale responses to respondent’s personal anchors
  – See: Gary King’s website on anchoring vignettes: http://gking.harvard.edu/vign/ (King et al., 2004; King & Wand, 2007)

• Growing in popularity
  – Used in surveys (e.g., sociology, political science, to measure health, SES)
Below you will find descriptions of three mathematics teachers. Read each of the descriptions of these teachers. Then let us know to what extent you agree with the final statement.

(Please check only one box on each row.)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Ms. Anderson assigns mathematics homework every other day. She always gets the answers back to students before examinations. <strong>Ms. Anderson is concerned about her students’ learning.</strong></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>b) Mr. Crawford assigns mathematics homework once a week. He always gets the answers back to students before examinations. <strong>Mr. Crawford is concerned about his students’ learning.</strong></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>c) Ms. Dalton assigns mathematics homework once a week. She never gets the answers back to students before examinations. <strong>Ms. Dalton is concerned about her students’ learning.</strong></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Below you will find descriptions of three mathematics teachers. Read each of the descriptions of these teachers. Then let us know to what extent you agree with the final statement.

(Please check only one box on each row.)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a)</strong> Ms. Anderson assigns mathematics homework every other day. She always gets the answers back to students before examinations.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Ms. Anderson is concerned about her students’ learning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>b)</strong> Mr. Crawford assigns mathematics homework once a week. He always gets the answers back to students before examinations.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Mr. Crawford is concerned about his students’ learning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>c)</strong> Ms. Dalton assigns mathematics homework once a week. She never gets the answers back to students before examinations.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Ms. Dalton is concerned about her students’ learning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Below you will find descriptions of three mathematics teachers. Read each of the descriptions of these teachers. Then let us know to what extent you agree with the final statement.

**Student “B’s” responses**

(Please check only one box on each row.)

<table>
<thead>
<tr>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

a) Ms. Anderson assigns mathematics homework every other day. She always gets the answers back to students before examinations. **Ms. Anderson is concerned about her students’ learning.**

b) Mr. Crawford assigns mathematics homework once a week. He always gets the answers back to students before examinations. **Mr. Crawford is concerned about his students’ learning.**

c) Ms. Dalton assigns mathematics homework once a week. She never gets the answers back to students before examinations. **Ms. Dalton is concerned about her students’ learning.**
Below you will find descriptions of three mathematics teachers. Read each of the descriptions of these teachers. Then let us know to what extent you agree with the final statement.

(Please check only one box on each row.)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Ms. Anderson assigns mathematics homework every other day. She always gets the answers back to students before examinations. Ms. Anderson is concerned about her students’ learning.</td>
<td>✗1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>b) Mr. Crawford assigns mathematics homework once a week. He always gets the answers back to students before examinations. Mr. Crawford is concerned about his students’ learning.</td>
<td>1</td>
<td>✗2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>c) Ms. Dalton assigns mathematics homework once a week. She never gets the answers back to students before examinations. Ms. Dalton is concerned about her students’ learning.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>✗4</td>
</tr>
</tbody>
</table>

My teacher lets students know they need to work hard.

Square 1 or 2

For Student “A” this can be interpreted as “like the middle hypothetical teacher”
Below you will find descriptions of three mathematics teachers. Read each of the descriptions of these teachers. Then let us know to what extent you agree with the final statement.

(Please check only one box on each row.)

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Ms. Anderson assigns mathematics homework every other day. She always gets the answers back to students before examinations. <strong>Ms. Anderson is concerned about her students’ learning.</strong></td>
<td>[ ] 1</td>
<td>[ ] 2</td>
<td>[X] 3</td>
<td>[ ] 4</td>
</tr>
<tr>
<td>b) Mr. Crawford assigns mathematics homework once a week. He always gets the answers back to students before examinations. <strong>Mr. Crawford is concerned about his students’ learning.</strong></td>
<td>[ ] 1</td>
<td>[ ] 2</td>
<td>[ ] 3</td>
<td>[X] 4</td>
</tr>
<tr>
<td>c) Ms. Dalton assigns mathematics homework once a week. She never gets the answers back to students before examinations. <strong>Ms. Dalton is concerned about her students’ learning.</strong></td>
<td>[ ] 1</td>
<td>[ ] 2</td>
<td>[ ] 3</td>
<td>[X] 4</td>
</tr>
</tbody>
</table>

My teacher lets students know they need to work hard. | [ ] 1 | [X] 2 | [ ] 3 | [ ] 4 |

For Student “B” this can be interpreted as “better than the best hypothetical teacher”
Correlations with Mathematics Achievement Scores

Country-level correlation (60 countries)

Average correlation within country (N = 1,000+ students)

Math self efficacy

Failure attribution

Disciplinary climate

Perceived control

Familiarity with math concepts

Math concept

Math interest

Weak anchors, abstract, vague

More context, concrete, countable

Personality

Math Anxiety

Correlations with Mathematics Achievement Scores

Country-level correlation (60 countries)
Correlations with Mathematics Achievement Scores

Average correlation within country ($N = 1,000+$ students)

Country-level correlation (60 countries)

-0.80 -0.60 -0.40 -0.20 0.00 +0.20 +0.40 +0.60

Within-country alignment of within-country and country-level correlations

Teacher-Support (anchored)

Mathematics Interest (anchored from Teacher-Support)

Teacher-Support (Likert)

Mathematics Interest (Likert)

Correlations with Mathematics Achievement Scores

Copyright © 2014 Educational Testing Service. All rights reserved.
Findings so far

- We have developed anchoring vignettes for many constructs, from Big 5 to emotional intelligence, for students and teachers
- Anchoring vignettes work very well on poorly anchored scales (much of personality assessment)
- They improve cross-country comparability; they also increase validity within a country
- Anchoring vignettes developed for one scale can be used to adjust other scales
- It is important to write vignettes so that students rate them appropriately
- Even without rescoring, they work by giving a frame of reference (if given before self ratings) (Hopkins & Wand, 2010)
<table>
<thead>
<tr>
<th>Does this method(address this problem)</th>
<th>Anchoring vignettes</th>
<th>Forced choice</th>
<th>Others’ ratings</th>
<th>Situational Judgment Tests (SJTs)</th>
<th>Behavior anchored rating scales (BARS)</th>
<th>Performance measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differentiation</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>sometimes</td>
<td>In principle</td>
</tr>
<tr>
<td>Cultural comparability</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>sometimes</td>
<td>probably</td>
</tr>
<tr>
<td>Social desirability</td>
<td>sometimes</td>
<td>yes</td>
<td>yes</td>
<td>somewhat</td>
<td>no</td>
<td>probably</td>
</tr>
<tr>
<td>Reference bias</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>In principle</td>
</tr>
<tr>
<td>Response style bias</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>In principle</td>
</tr>
</tbody>
</table>

**Disadvantages**

<table>
<thead>
<tr>
<th></th>
<th>More testing time</th>
<th>High dev’Ip costs</th>
<th>Not ready for use</th>
<th>3^rd party involved</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>xx</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>High dev’Ip costs</td>
<td></td>
<td>xx</td>
<td>xx</td>
<td>?</td>
</tr>
<tr>
<td>Not ready for use</td>
<td></td>
<td></td>
<td></td>
<td>xxx</td>
</tr>
<tr>
<td>3^rd party involved</td>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
</tbody>
</table>
Summary

- **Anchoring vignettes** can increase validity and address cross-cultural comparability; they require more time.
- **Forced-choice** additionally prevent socially desirable responses (“faking good”) but take even more time.
- **Others’ (teachers, parents) ratings** are useful in that they provide a different frame of reference, and research shows they are more predictive of outcomes.
- Other methods (**SJT**, **BARS**) are useful; they have high development costs.
- **Performance measures** are potentially ideal but we do not have many of these, yet.
- Using one or more of these techniques can increase the quality of noncognitive data available for use in analysis & policy.


References


