

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 Paulo, Brazil
March 25, 2014**

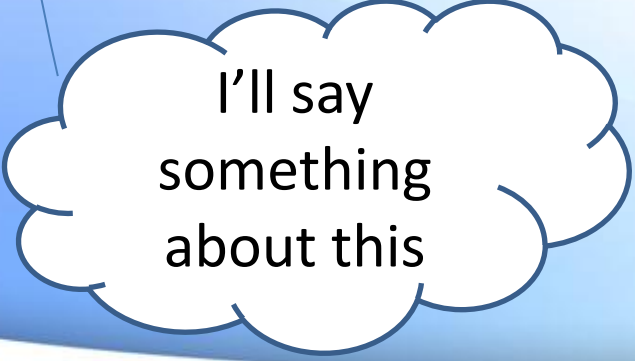


Why alternative measures for noncognitive assessment?

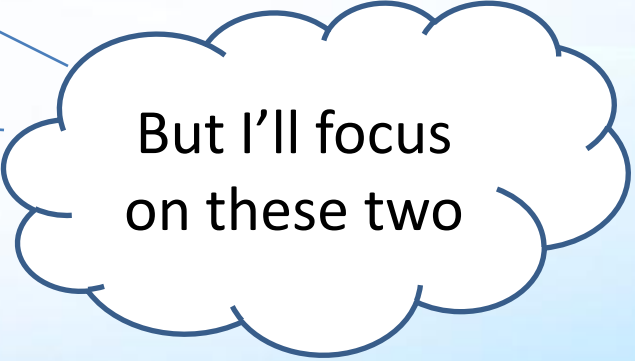
- Self-rating scales are very useful as is—most of what we know about noncognitive skills is based on such scales!
- But there are problems with these scales
 - Socially desirable responding (wanting to look good)
 - Reference group bias (who you compare to)
 - Response style bias (e.g., extreme responses, 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**



I'll say something about this



But I'll focus on these two

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

		Strongly disagree	Disagree	Agree	Strongly agree
1.	I keep my promises	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2.	I am generally pretty forgiving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Forced-Choice

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

1.	I keep my promises	<input type="radio"/>
2.	I am generally pretty forgiving	<input type="radio"/>

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

Country-level correlations (n = 19) between		UN Human Development Index (education, life expectancy, GDP)	Global competitive index (WEF), requirements, efficiency, innovation
Agreeableness	Single Statement	.09	.39
	Forced Choice	.57	.58
Emotional stability	Single Statement	.07	.50
	Forced choice	.27	.53
Extraversion	Single Statement	.41	.20
	Forced choice	.76	.46
Conscientiousness	Single Statement	-.46	-.40
	Forced Choice	-.08	.21

(Bartram, 2013)

ANCHORING VIGNETTES



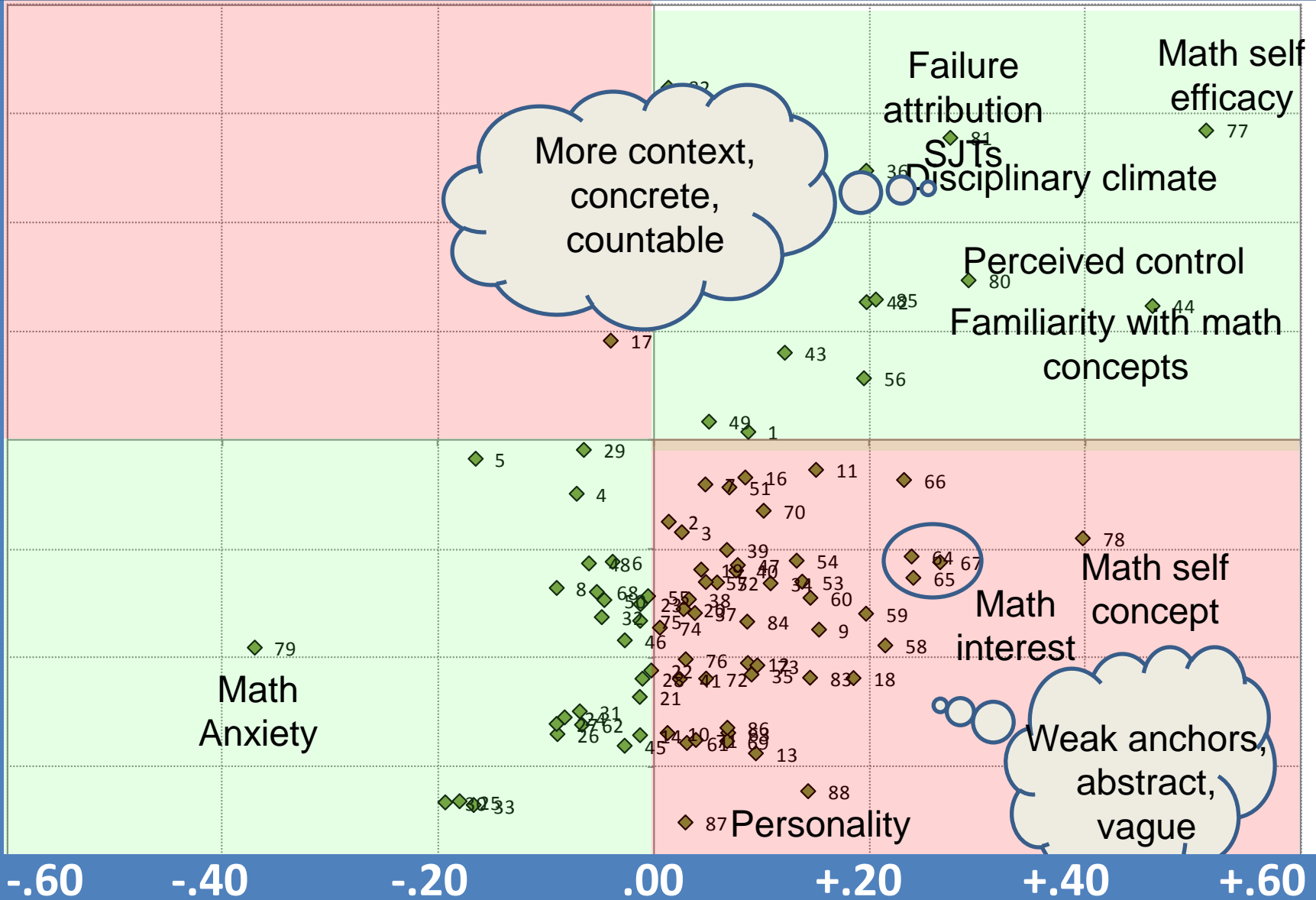
Cross-Cultural Validity

- 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

Country-level correlation (60 countries)

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



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)

01 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.)

	<i>Strongly agree</i>	<i>Agree</i>	<i>Disagree</i>	<i>Strongly disagree</i>
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.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
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.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
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.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄

01 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 "A's" responses

(Please check only one box on each row.)

Strongly agree Agree Disagree Strongly disagree

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.**

₁ ₂ ₃ ₄

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.**

₁ ₂ ₃ ₄

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.**

₁ ₂ ₃ ₄

01 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.)

<i>Strongly agree</i>	<i>Agree</i>	<i>Disagree</i>	<i>Strongly disagree</i>
-----------------------	--------------	-----------------	--------------------------

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.**

<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input checked="" type="checkbox"/> ₃	<input type="checkbox"/> ₄
---------------------------------------	---------------------------------------	--	---------------------------------------

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.**

<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input checked="" type="checkbox"/> ₄
---------------------------------------	---------------------------------------	---------------------------------------	--

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.**

<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input checked="" type="checkbox"/> ₄
---------------------------------------	---------------------------------------	---------------------------------------	--

01 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 "A's" responses

(Please check only one box on each row.)

- | | <i>Strongly agree</i> | <i>Agree</i> | <i>Disagree</i> | <i>Strongly disagree</i> |
|--|--|--|---------------------------------------|--|
| 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. | <input checked="" type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input type="checkbox"/> ₄ |
| 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. | <input type="checkbox"/> ₁ | <input checked="" type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input type="checkbox"/> ₄ |
| 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. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input checked="" type="checkbox"/> ₄ |

02 My teacher lets students know they need to work hard.

<input type="checkbox"/> ₁	<input checked="" type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
---------------------------------------	--	---------------------------------------	---------------------------------------

For Student "A" this can be interpreted as "like the middle hypothetical teacher"

01 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.)

- | | <i>Strongly agree</i> | <i>Agree</i> | <i>Disagree</i> | <i>Strongly disagree</i> |
|--|---------------------------------------|---------------------------------------|--|--|
| 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. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input checked="" type="checkbox"/> ₃ | <input type="checkbox"/> ₄ |
| 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. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input checked="" type="checkbox"/> ₄ |
| 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. | <input type="checkbox"/> ₁ | <input type="checkbox"/> ₂ | <input type="checkbox"/> ₃ | <input checked="" type="checkbox"/> ₄ |

02 My teacher lets students know they need to work hard.

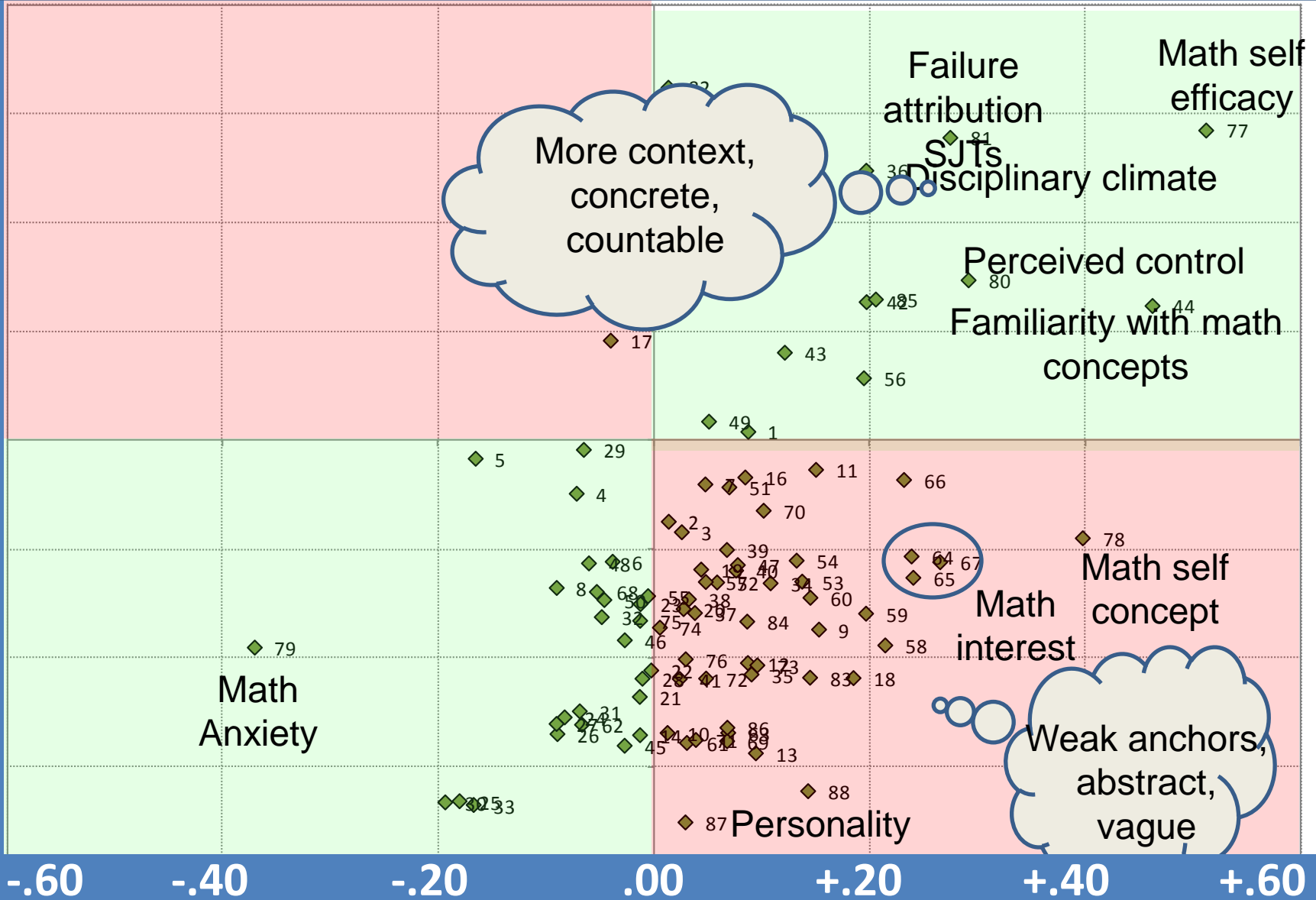
<input type="checkbox"/> ₁	<input checked="" type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
---------------------------------------	--	---------------------------------------	---------------------------------------

For Student "B" this can be interpreted as "better than the best hypothetical teacher"

Correlations with Mathematics Achievement Scores

Country-level correlation (60 countries)

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

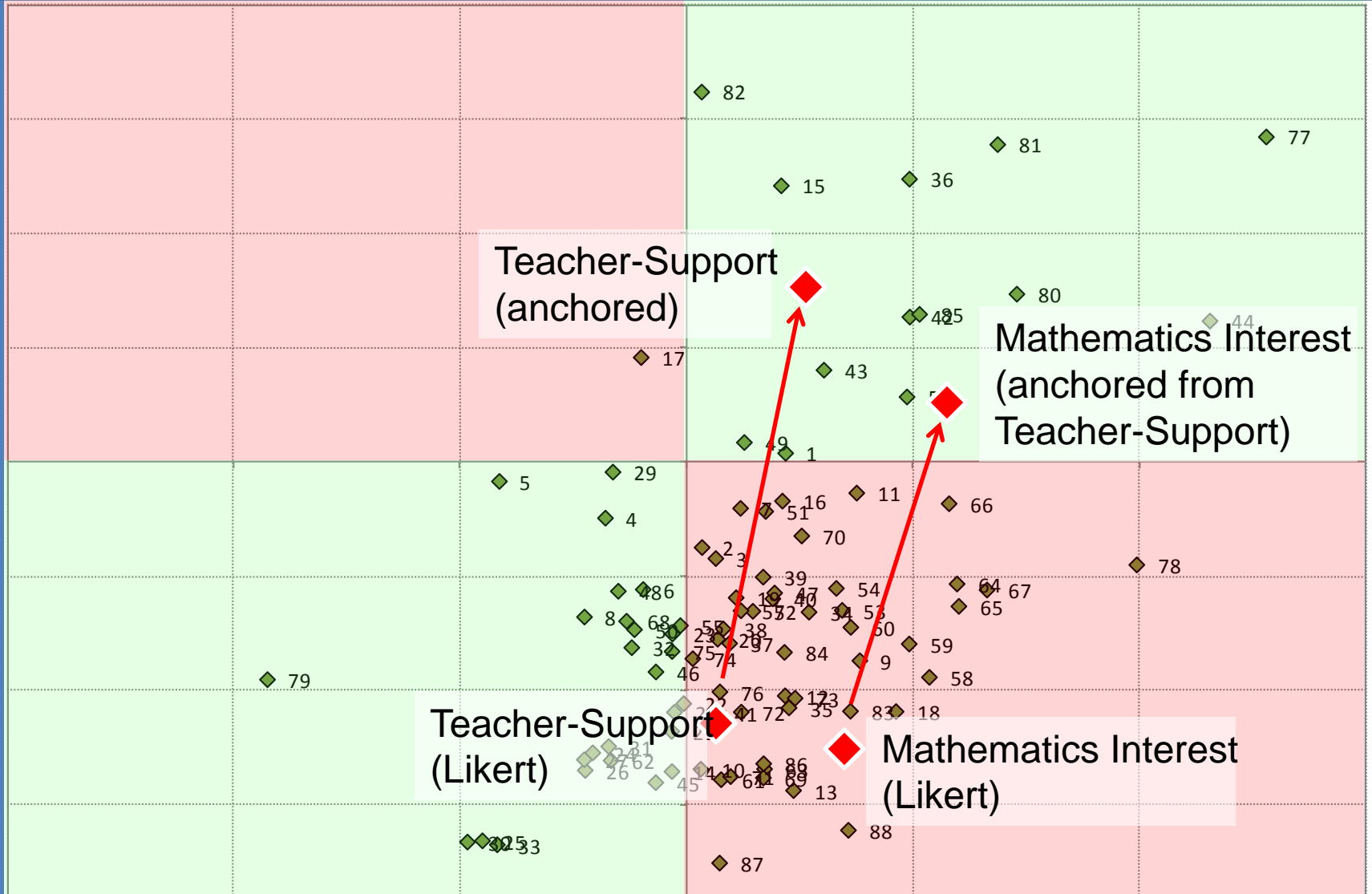


Average correlation within country (N = 1,000+ students)

Correlations with Mathematics Achievement Scores

Country-level correlation (60 countries)

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



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

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)

SUMMARY



Does this method → address this ↓ problem ?	Anchoring vignettes	Forced choice	Others' ratings	Situational Judgment Tests (SJTs)	Behavior anchored rating scales (BARS)	Per- formance measures
Differentiation	yes	yes	no	yes	sometimes	In principle
Cultural comparability	yes	yes	no	no	sometimes	probably
Social desirability	sometimes	yes	yes	somewhat	no	probably
Reference bias	yes	yes	no	yes	yes	In principle
Response style bias	yes	yes	no	yes	yes	In principle
Disadvantages	↓	↓	↓	↓	↓	↓
More testing time	X	XX		XX		
High dev'lp costs				XX	XX	?
Not ready for use						XXX
3 rd party involved			X		X	?

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 (**SJTs, 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

- Bartram, D. (2013). Scalar equivalence of OPQ32: Big Five profiles of 31 countries. *Journal of Cross-Cultural Psychology, 44*(1), 61-83.
- Brown, A., & Bartram, D. (2009, April). *Doing less but getting more: Improving forced-choice measures with IRT*. Paper presented at the 24th annual conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- Brown, A. & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods, 18*(1), 36-52. DOI: 10.1037/a0030641
- Connelly BS & Ones DS (2010). An other perspective on personality: meta-analytic integration of observers' accuracy and predictive validity. *Psychol Bull, 136*(6), 1092-1122.
- Dragow, F., Stark, S., Chernyshenko, O.S., Nye, C.D., Hulin, C.L., White, L. A. (2012). *Development of the Tailored Adaptive Personality Assessment System (TAPAS) to Support Army Selection and Classification Decisions*. ARI Technical Report 1311. Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences
- Hopkins, D., & King, G. (2010). Improving Anchoring Vignettes: Designing Surveys to Correct Interpersonal Incomparability. *Public Opinion Quarterly: 1-22*.

References

- King, G., Christopher JL ,Murray, C.J.L., Salomon, J.A., & Tandon, A. (2004). Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research. *American Political Science Review* 98: 191–207
- King, G., & Wand, J. (2007). Comparing Incomparable Survey Responses: New Tools for Anchoring Vignettes. *Political Analysis* 15: 46-66.
- Kyllonen, P.C. (2013). Soft skills for the workplace. *Change*, 45(6), 16-23.
- Kyllonen, P.C., & Bertling, J. (2014). Innovative questionnaire assessment methods to increase cross-country comparability. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*. Boca Raton, FL: CRC Press.
- Lindqvist, E., Vestman, R. (2011). The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment. *American Economic Journal: Applied Economics*, 3(1): 101-28.
- Oh, I.-S., Wang, G., Mount, M.K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, 96(4), 762-773.
- Salgado, J.F., & Táuriz, G. (2012): The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies, *European Journal of Work and Organizational Psychology*, DOI:10.1080/ 1359432X.2012.716198
- Segal, C. (2013). Misbehavior, Education, and Labor Market Outcomes, *The Journal of the European Economic Association*, 11(4), 743-779.