

Are self-assessments of cognitive and non-cognitive development reliable?

Michael Roona

UC Merced

Types of assessment instruments

- “Objective” tests
 - Narrowly focused
 - Wide substantive coverage requires multiple tests
 - Extensive testing consumes class time
 - Expensive
- “Subjective” reflections
 - Can be expansive
 - Breadth and depth can be assessed with a single instrument
 - Gathering survey data requires little class time
 - Inexpensive

20 UCUES learning outcome measures

- Critical thinking skills
- Writing skills
- Reading comprehension
- Foreign language skills
- Disciplinary knowledge
- Quantitative skills
- Speaking skills (English)
- International understanding
- Leadership skills
- Computer skills
- Internet skills
- Library research skills
- Other research skills
- Presentation skills
- Interpersonal (social) skills
- Appreciation, tolerance of racial and ethnic diversity
- Appreciation of fine arts
- Appreciation of cultural and global diversity
- Understanding import of personal responsibility
- Self-awareness and understanding

How UCUES measures learning

Academic and Personal Development

2. Please rate your level of proficiency in the following areas when you started at this campus and now.

	When you started here	Current ability level
Analytical and critical thinking skills	<input type="radio"/> Very poor	<input type="radio"/> Very poor
	<input type="radio"/> Poor	<input type="radio"/> Poor
	<input type="radio"/> Fair	<input type="radio"/> Fair
	<input type="radio"/> Good	<input type="radio"/> Good
	<input type="radio"/> Very good	<input type="radio"/> Very good
	<input type="radio"/> Excellent	<input type="radio"/> Excellent

Evidence about self-report validity is mixed

- Self-reports of knowledge or ability generally correlate well with more objective measures
- Self-reported GAINS in ability or knowledge are subject to well documented conscious and subconscious biases (e.g., Halo effect)
- “Can first-year college students provide accurate self-reports about their learning and development?” by Nicholas A. Bowman includes an excellent lit review

“Scientific” studies of validity

- Assume “objective” test scores are the gold standard (since Thorndike beat Dewey)
- Assume parametric inferential statistics are “more rigorous” than non-parametric or descriptive statistics
 - Nearly all studies contrasting objective test scores with subjective self-reports use parametric statistical tests, even though most self-reports involve ordinal variables

Study Design and Methods

- First time freshmen in Fall 2005 who
 - completed UCUES in Spring 2006 and
 - completed UCUES in Spring 2008 (N=94)
- Compare “when you started here” ratings provided as freshmen and as juniors by
 - using Goodman & Kruskal’s Gamma
 - using Cohen’s Kappa (controversial)
 - examining distributions of differences between freshman and junior ratings of baseline abilities

Meaning of ordinal (non-parametric) tests

- Goodman and Kruskal's Gamma
 - measures concordance (highest value is 1)
 - $\text{Gamma} = 1$ if all observations fall on the diagonal from upper left to lower right of cross-tabulation, i.e., perfect relationship between pre & post tests
- Cohen's Kappa
 - measures proportion of agreements after chance agreements (expected values) are excluded
 - should only be used with independent samples, but is often used for inter-rater reliability analyses

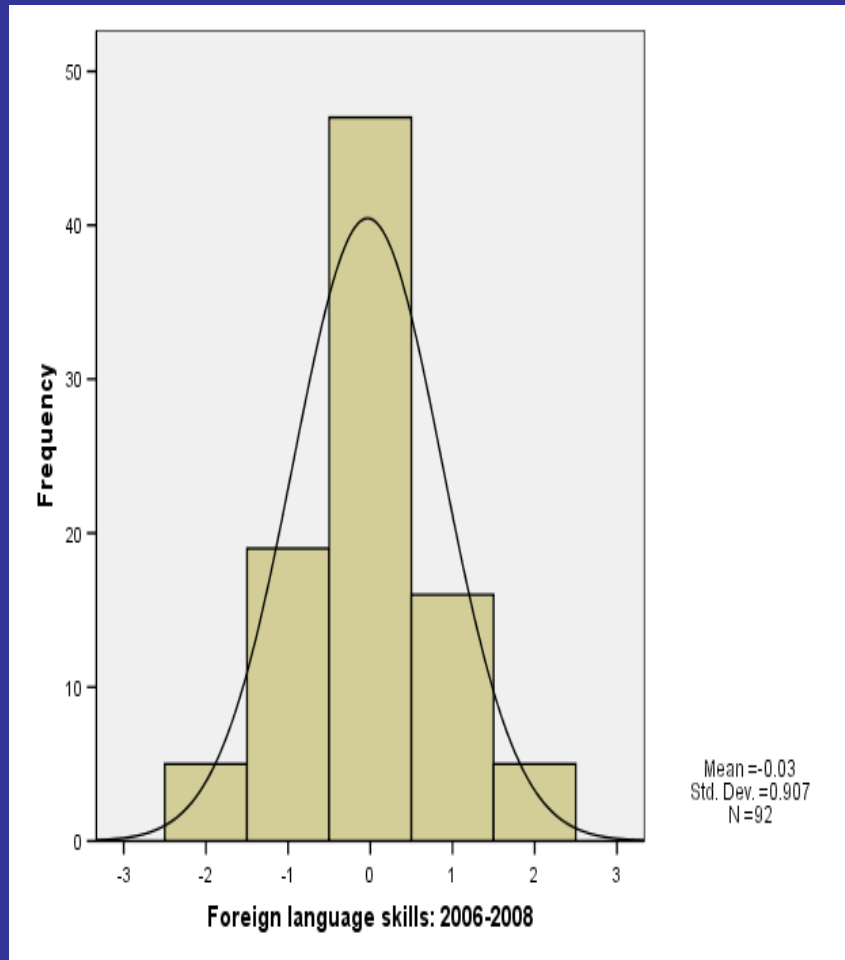
Concordance on “when you started here”

	Gamma	Sig	Bonf
ability to be clear and effective when writing	0.781	0.000	
quantitative (mathematical and statistical) skills	0.768	0.000	
foreign language skills	0.730	0.000	
ability to appreciate cultural and global diversity	0.701	0.000	
internet skills	0.680	0.000	
analytical and critical thinking skills	0.656	0.000	
ability to read and comprehend academic material	0.634	0.000	
library research skills	0.623	0.000	
ability to appreciate the fine arts	0.610	0.000	
interpersonal (social) skills	0.591	0.000	
leadership skills	0.575	0.000	
ability to speak clearly and effectively in English	0.567	0.000	
computer skills	0.541	0.000	
ability to appreciate racial and ethnic diversity	0.528	0.000	
other research skills	0.469	0.000	
ability to prepare and make a presentation	0.460	0.000	
understanding international perspectives	0.390	0.001	
understanding of a specific field of study	0.350	0.006	0.120
understanding the import of personal responsibility	0.348	0.005	0.100
self awareness and understanding	0.329	0.003	0.060

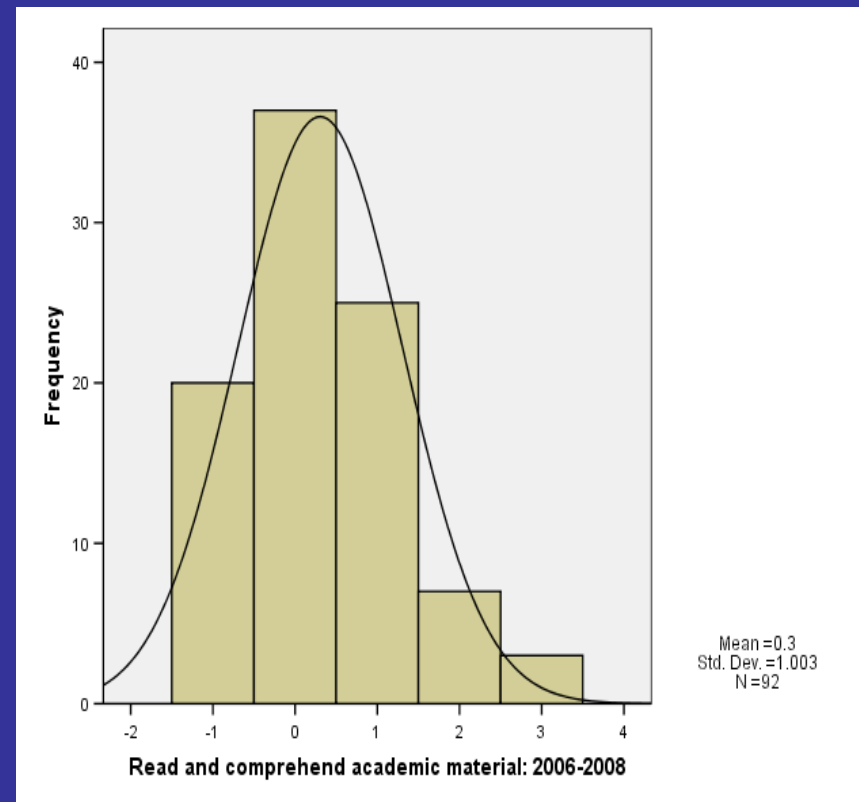
Distribution of 06 minus 08 self-report diffs

	Mean	St Dev	Skew	Std Err
foreign language skills	-0.033	0.907	0.065	0.251
ability to appreciate racial and ethnic diversity	0.086	1.060	0.273	0.250
ability to be clear and effective when writing	0.098	0.742	0.171	0.251
analytical and critical thinking skills	0.098	0.878	-0.094	0.251
library research skills	0.118	1.009	-0.242	0.250
ability to appreciate cultural and global diversity	0.140	0.928	0.133	0.250
computer skills	0.163	1.179	-0.406	0.251
understanding the import of personal responsibility	0.183	1.197	-0.012	0.250
ability to appreciate the fine arts	0.213	1.025	-0.380	0.249
leadership skills	0.217	1.098	-0.496	0.251
understanding of a specific field of study	0.228	1.159	0.146	0.251
ability to speak clearly and effectively in English	0.272	0.996	-0.232	0.251
quantitative (mathematical and statistical) skills	0.304	0.861	-0.109	0.251
ability to read & comprehend academic material	0.304	1.003	0.623	0.251
understanding international perspectives	0.330	1.265	0.158	0.253
ability to prepare and make a presentation	0.337	1.102	-0.204	0.251
interpersonal (social) skills	0.337	1.141	-0.290	0.251
other research skills	0.348	1.104	0.171	0.251
internet skills	0.413	0.974	-0.042	0.251
self awareness and understanding	0.473	1.315	-0.380	0.250

Symmetric “error”



Positively skewed



Distribution of 06 minus 08 self-report diffs

	Mean	St Dev	Skew	Std Err
foreign language skills	-0.033	0.907	0.065	0.251
ability to appreciate racial and ethnic diversity	0.086	1.060	0.273	0.250
ability to be clear and effective when writing	0.098	0.742	0.171	0.251
analytical and critical thinking skills	0.098	0.878	-0.094	0.251
library research skills	0.118	1.009	-0.242	0.250
ability to appreciate cultural and global diversity	0.140	0.928	0.133	0.250
computer skills	0.163	1.179	-0.406	0.251
understanding the import of personal responsibility	0.183	1.197	-0.012	0.250
ability to appreciate the fine arts	0.213	1.025	-0.380	0.249
leadership skills	0.217	1.098	-0.496	0.251
understanding of a specific field of study	0.228	1.159	0.146	0.251
ability to speak clearly and effectively in English	0.272	0.996	-0.232	0.251
quantitative (mathematical and statistical) skills	0.304	0.861	-0.109	0.251
ability to read & comprehend academic material	0.304	1.003	0.623	0.251
understanding international perspectives	0.330	1.265	0.158	0.253
ability to prepare and make a presentation	0.337	1.102	-0.204	0.251
interpersonal (social) skills	0.337	1.141	-0.290	0.251
other research skills	0.348	1.104	0.171	0.251
internet skills	0.413	0.974	-0.042	0.251
self awareness and understanding	0.473	1.315	-0.380	0.250

Findings

- Goodman and Kruskal's Gamma indicates:
 - concordant self-ratings of baseline ability two years apart for basic skills that are routinely used and/or tested, and hence clearly defined (writing, math, language, internet)
 - discordant self-ratings regarding more ambiguous concept of level of understanding (of self or of an academic field)
- Difference distribution statistics indicate:
 - on average, students assess their ability when they entered college higher when they were freshmen than juniors
 - these differences are not a result of positive skewness; they result from positive shift of entire distribution

Tentative conclusions

- “Subjective” measures like self-assessments are:
 - “reliable” when the ability being measured is unambiguous
 - prone to measurement error when ambiguity is introduced by the learning process (e.g., when “specific field of study” changes from “biology” for freshmen to “cellular and molecular biology” for juniors)
- “Subjective” measures like self-assessments may be useful for assessing student learning, but
 - must be captured on the same instrument to control for measurement error
 - should not be used in pretest-posttest or cross-sectional studies
- These inferences contradict the received view that because people make better judgments about their current abilities than their past abilities, retrospective judgments are suspect

“No man ever steps in the same river twice, for it's not the same river and he's not the same man.”

- Heraclitus

