



Transparent Quality: Framing and building a Psychometric Portfolio

Justin Paulsen and Allison BrckaLorenz
Indiana University
Center for Postsecondary Research

Program presented at the 2018 Association for Institutional Research Forum, Orlando, FL

Agenda

- Why build a portfolio for data quality?
- Quick overview of NSSE and FSSE
- Explore major frameworks for operationalizing quality (validity, reliability, data quality)
- Definitions, examples, and NSSE/FSSE studies
- An example portfolio
- Discuss challenges
- Final thoughts and questions



Why are you interested in transparent quality?

Why We're Here

- Data is powerful, and it's important to us to build and maintain trust with the varied audiences that use our data and results
- Sometimes people want to interrogate the data instead of have conversations about the quality of undergraduate education
- People have legitimate fears about bias, misrepresentation, and other dangers associated with the "average" student
- "Well...you can't trust self-reported data"
- Allison loves small population analysis—the validity of such analyses is commonly questioned!

NSSE and FSSE

- National Survey of Student Engagement (NSSE)
 - Annual survey of first-year and senior students at four-year colleges and universities
 - NSSE asks students about the time and effort students put into behaviors that have been linked to student learning and development
 - In 2017: 500,000+ respondents at 725 institutions
- Faculty Survey of Student Engagement (FSSE)
 - Annual survey of faculty and instructional staff who teach at least one undergraduate course in the academic year of administration
 - Measures faculty expectations and values for student engagement and their use of educational practices linked with high levels of learning and development
 - In 2017: 24,000+ respondents at 154 institutions

Data Quality Frameworks

- Borsboom, Mellenbergh, van Heerden (2004) – Causal Validity
- Lissitz & Samuelson (2007) – Content Validity
- Messick (1989) – Unified Construct of Validity

Causal Validity

(Bonboom et al., 2004)

Construct



- Validity = Variation in the construct results in variations in the responses to the instrument
- Validity is something you put into an instrument
- Design items based on the processes necessary to answer the instrument item correctly (example: Cognitive Diagnostic Models)
- Difficult to prove, rarely used

Instrument



Content Validity

(Lissitz & Samuelson, 2007)

		Perspective		
		Theoretical	Practical	
Investigative Focus	Internal	Latent Process	Content and Reliability	<ul style="list-style-type: none"> a) Test Blueprint b) Item difficulty, discrimination c) DIF d) Reliability e) Rater Consistency
	External	Nomological Network	Utility and Impact	
		<ul style="list-style-type: none"> a) Factor Analysis b) Convergent-Divergent Correlation c) Think-Alouds d) Item Intercorrelation 		
		<ul style="list-style-type: none"> a) Multitrait Multimethod Matrix b) Confirmatory Factor Analysis c) Analysis of Variance d) SEM 	<ul style="list-style-type: none"> a) Regression Analysis b) Cost-Benefit Analysis c) Examine decisions d) Evaluation of test use 	

Unified Validity

(Messick, 1989; Goodwin & Leach, 2003; Clark, 2016)

Unified Validity				
Evidence: Test Content	Evidence: Response Processes	Evidence: Internal Structure	Evidence: Relations to Other Variables	Evidence: Consequences of Testing
<ul style="list-style-type: none"> • Subject Matter Experts' Analysis of Items • Investigation of construct irrelevance or construct under-representation 	<ul style="list-style-type: none"> • Test Taker Talk-Alouds • Process Data • Rater Uniformity 	<ul style="list-style-type: none"> • Confirmatory Factor Analysis • Inter-Item correlation • Internal Consistency • Differential Item Functioning 	<ul style="list-style-type: none"> • Criterion-Related Validity • Convergent-Divergent Correlation • Intervention-Experiment Study 	<ul style="list-style-type: none"> • Studies of Anticipated Benefits • Studies of Negative Consequences • Cost-Benefit Analysis • Evaluation of differential impacts by group

Evidence Based on Test Content

What is it?

- Evidence used to show that the content of the survey is representative of the content domain

Examples

- Subject matter experts' analysis of items
- Investigation of construct irrelevance or construct under-representation
- Following the guidance of conceptual frameworks or association standards

NSSE Example: Test Content Evidence

- Validating Civic Engagement Topical Module
 - Summarize literature (Ehrlich, Jacoby)
 - Ask content experts to review items
 - Are we measuring what is important?
 - Are these items actionable?
 - Do you recommend additional items?
 - Examine existing association standards (e.g. Institute for Democracy & Higher Education, Center for Information and Research on Civic Learning and Engagement)

Evidence Based on Response Process

What is it?

- Evidence that indicates the relationship between the construct and the actual performance on the item

Examples

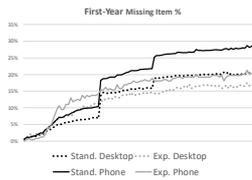
- Test Taker Talk-Alouds
- Process Data
- Rater Uniformity

NSSE Example: Response Process Evidence

Cognitive Interviews

- "Tell me how you determined your answer"
- "What does synthesizing mean in 'Synthesizing an idea, experience, or line of reasoning in depth'"
 - Finding: 'Synthesizing' and 'Analyzing' interpreted the same

Process Data



Evidence Based on Internal Structure

What is it?

- Evidence indicating the degree to which items relate to the conceptual framework

Examples

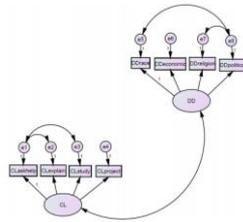
- Confirmatory Factor Analysis
- Inter-item correlation
- Internal Consistency
- Differential Item Functioning

NSSE Example: Internal Structure Evidence

Table 3a. Experiences with Faculty: CFA Model Fit Indices

	All Seniors	All First-Year Students	Online Seniors	Online First-Year Students
N	89,391	99,976	10,296	2,488
CMIN/DF	183.533	98.421	30.266	8.357
GFI	.993	.995	.990	.988
CFI	.993	.993	.989	.989
RMSEA	.045	.040	.053	.055
PCLOSE	1.00	1.00	.094	.174

GFI: .85 or higher
CFI: .90 or higher
RMSEA: .06 or lower
PCLOSE: .05 or higher



Analysis conducted using AMOS, but can also be conducted in STATA, R, Mplus

FSSE Example: Internal Structure Evidence

Table 2. Scale Cronbach's Alphas by Course Division

FSSE scales	Cronbach's α		Inter-Item Correlation		Average Inter-Item Correlation	
	Lower Division	Upper Division	Lower Division	Upper Division	Lower Division	Upper Division
Higher-Order Learning	.74	.75	.33-.63	.29-.58	.42	.39
Reflective & Integrative Learning	.89	.87	.39-.78	.29-.76	.53	.49
Learning Strategies	.76	.77	.40-.62	.48-.62	.52	.53
Quantitative Reasoning	.88	.87	.64-.80	.61-.77	.71	.68
Collaborative Learning	.85	.83	.55-.78	.46-.76	.60	.56
Discussions with Others	.84	.83	.22-.84	.29-.83	.79	.78
Student-Faculty Interaction	.77	.76	.35-.53	.34-.52	.47	.47
Effective Teaching Practices	.76	.77	.24-.42	.26-.45	.30	.30
Quality of Instruction	.83	.80	.38-.73	.45-.74	.53	.54
Supportive Environment	.87	.86	.29-.66	.29-.62	.44	.44

Table 3. Scale Item Scale Analysis by Course Division

FSSE scales	Cronbach's α if Item Deleted				Corrected Item-Scale Correlation
	Lower Division	Upper Division	Lower Division	Upper Division	
Higher-Order Learning	.61-.80	.61-.77	.29-.62	.31-.61	
Reflective & Integrative Learning	.86-.89	.83-.88	.49-.76	.41-.76	
Learning Strategies	.59-.76	.62-.78	.51-.67	.53-.67	
Quantitative Reasoning	.78-.89	.72-.89	.70-.82	.68-.80	
Collaborative Learning	.79-.80	.76-.85	.58-.77	.55-.75	
Discussions with Others	.75-.82	.80-.82	.62-.82	.62-.82	
Student-Faculty Interaction	.68-.75	.69-.75	.53-.64	.53-.65	
Effective Teaching Practices	.73-.75	.72-.75	.41-.53	.42-.55	
Quality of Instruction	.81-.86	.81-.85	.52-.73	.55-.74	
Supportive Environment	.84-.86	.84-.86	.53-.69	.53-.70	

Table 1. Internal Consistency Criteria for This Study

Reliability Statistics	Criteria for a Good Scale
Cronbach's Alpha	Greater than or equal to .70
Range of inter-item correlations	Between .15 and .35
Average inter-item correlation	Between .35 and .50
Range of Cronbach's alpha's if item deleted	Dropping any item should decrease the alpha
Range of corrected item-scale correlations	Greater than or equal .50

Analysis conducted using "Reliability" command in SPSS

FSSE Example: Internal Structure Evidence

All Items

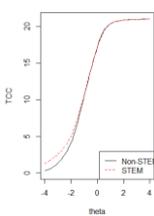


Table 2. DIF in the Reflective & Integrative Learning Scale

	STEM	Division	Online	Adjunct	Full-Time	Gender	White
RIntegrate	0.03	0.03		0.003	0.001		0.001
RIntocultural			0.001	0.001	0.001	0.001	0.001
RIntolanguage			0.001			0.001	0.001
RIntoreview			0.001				
RIntospect			0.001	0.001			
RIntoreview				0.001	0.001		0.001
RIntconnect	0.01			0.001	0.001	0.001	0.001

Analysis conducted using "lordif" package in R

Evidence Based on Relations to Other Variables

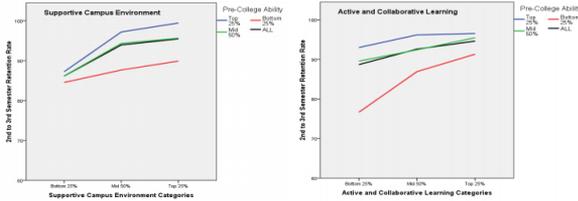
What is it?

- Evidence indicating the degree to which relationships are consistent with the hypothesized construct

Examples

- Criterion-Related Validity
- Convergent-Divergent Correlation
- Intervention-Experiment Study

NSSE Example: Relation to Variables Evidence



Evidence Based on Consequences of Testing

What is it?

- Evidence indicating the degree to which consequences of test score use are related to the conceptual framework of the test

Examples

- Studies of Anticipated Benefits
- Studies of Negative Consequences
- Cost-Benefit Analysis
- Evaluation of differential impacts by group

NSSE Example: Consequences of Testing

- University of Nebraska – Lincoln: *NSSE findings called attention to the need to revisit UNL's learning outcomes and the structure of its general education program. UNL provided each college with a detailed report of their students' NSSE responses. Some colleges shared the results with other constituent groups (students, alumni, faculty members), and all colleges used the results as benchmark data.*
- University of Colorado, Boulder: *Results from four NSSE administrations on several survey items related to student writing showed that CU-Boulder students improved over time and compared well to their counterparts at peer institutions.*
- Texas Lutheran University: *Following each year of its participation in the 2000, 2003, and 2006 NSSE survey, Texas Lutheran has scheduled faculty workshops to discuss its students' responses to the survey. Discussions of the results after the first two administrations helped to increase faculty awareness of institutional strengths as well as areas of needed improvement.*

NSSE and FSSE's Portfolio

- NSSE and FSSE have a slightly more complicated scheme loosely based off Messick (1989)
- Validity
 - Response process validity
 - Content validity
 - Construct validity
 - Concurrent validity
 - Predictive validity
 - Known-groups validity
 - Consequential validity
- Reliability
 - Internal consistency
 - Temporal stability
 - Equivalence
- Other data quality indicators
 - Self-selection bias
 - Measurement error
 - Data quality (missing data)
 - Mode analysis
 - Nonresponse effects/bias
 - Sampling error
 - Social desirability

FSSE's Content Summaries

- To further transparency for FSSE findings, we started posted Content Summaries
- Item level frequencies/percentages, means, standard deviations, factor loadings (when appropriate)
- Scale-level means, standard deviations, Cronbach's α , ICC, correlations with other measures, distributions by disciplinary area, significant predictors by faculty, course, and institution characteristics



Challenges

- TIME
 - Rotate studies for longitudinal or cyclical assessments
 - Find yourself a "Justin"
- Knowledge of new techniques
 - Collaboration!
- That one audience member that just won't stop picking on the data/methods/survey items/etc.
 - This one is tough! But having a portfolio of data quality studies is very useful!

What are your greatest challenges?

What solutions have worked for you?

Recommended Studies

Short on time—start here!

Test content

- Get a quick expert opinion on 1) content and 2) structure

Internal structure

- Exploratory factor analysis
- Intercorrelations

Relations to other variables

- Convergent-divergent correlation

If you're using someone else's instrument (NSSE, FSSE, etc.), see what tests they have!

When you have more time

Response process

- Interviews or think aloud's

Internal structure

- Confirmatory factor analysis
- Differential item functioning

Relations to other variables

- Criterion-related studies

Consequences of testing

- How are people using your results?!

Final thoughts and questions?

Thanks for joining us!

Allison: abrcckalo@indiana.edu

Website: FSSE.indiana.edu

E-mail: FSSE@indiana.edu

Blog: NSSEsightings.indiana.edu

@NSSEsurvey  



References and Resources

- NSSE and FSSE Psychometric Portfolios
 - http://nsse.indiana.edu/html/psychometric_portfolio.cfm
 - http://fsse.indiana.edu/html/psychometric_portfolio.cfm
- References
 - Borsboom, Mellenbergh, van Heerden (2004)
 - Cizek (2016)
 - Goodwin & Leach (2003)
 - Lissitz & Samuelson (2007)
 - Messick (1989)