Development and Initial Testing of the Online Undergraduate Statistics Inventory

Raymond D. Mooring[1]
[1]Analysis Made Easy, 1941 Boulder Gate Drive, Ellenwood, GA 30294

Abstract
As technology advances, courses that were once taught using chalkboards in classrooms (e.g. *Introductory Statistics*) are now being offered online by various university departments. To be successful, the students in these courses must master the statistics content and navigate the online environment. Previous efforts to predict success in statistics have not included data from students in online sections. Because of this, the predictive power of skills germane to the online environment is not taken into consideration. On the other hand, inventories have been developed to predict academic success in the online environment using data from general courses, not STEM subjects. The inventories fail to capture the self-discipline and technical competence needed to succeed in a STEM subject.

This study begins to fill this void. The goal is to develop an inventory (questionnaire) that can predict success in a fully online undergraduate statistics class. The psychometric properties (calculated under the Rasch framework) of the first iteration of the Online Undergraduate Statistics Inventory (OUS-I) are presented here. The dimensionality of the OUS-I, effectiveness of the reverse coding used in the instrument, and the effectiveness of the rating scale are investigated.

Key Words: Introductory Statistics; Inventory; Psychometrics; STEM; Rasch Measurement; Distance Learning

1. Developing the Data Collection Instrument

1.1 Initial Stages

The development of the data collection instrument (also referred to as "questionnaire") began with the author creating an item bank of 64 statements designed to load onto 11 constructs that tap statistics and online education dimensions. Each of the 11 theoretical constructs was hypothesized to be related to success in a fully online undergraduate statistics course.

1.2 Face Validity

After the item bank was created, an email was sent to the members of the *Statistical Education* and *Statistical Consulting* sections of the American Statistical Association requesting volunteers interested in establishing the face validity of the statements. The email requested volunteers who had taught a fully online statistics course within the past two years. Five statistics instructors were selected from the email solicitation. A *Qualtrics* survey was designed to capture the volunteers' input.

Each volunteer was given the names of 11 hypothetical constructs (definitions for each construct were not provided) and was asked which construct s/he believed each of the 64 statements was attempting to tap. The reviewers could only select one construct for each statement so each reviewer had to independently decide which construct each was best aligned to.

It became clear early in the process that statements could potentially load onto multiple constructs considering each reviewer could develop his or her own definitions for the constructs. Accordingly, 24 statements were found to load onto three or more constructs. After removing these ambiguous statements, an item bank of 40 questions remained.

1.3 Survey Administration

Two of the five statistics instructors who had helped to identify the OUS-I's face validity were selected to assist in the second phase of its development: determination of its psychometric properties. In addition to the OUS-I author, these two instructors were scheduled to teach a fully online undergraduate statistics course in the spring and/or summer of 2015. At the beginning of each semester, we asked our students to complete the OUS-I questionnaire using the *Qualtrics* survey platform. The choice to incentivize completion was left to each instructor (one instructor elected to add bonus points to a test for completion).

1.3.1 Respondents

A total of 106 undergraduate students from four universities in the southern and southeastern region of the United States provided the initial data for the study. Demographic information was collected on each participant but was not analyzed.

Each respondent was asked to rate their level of agreement with each statement (on a 7-point scale). Several constructs contained statements that were reverse-coded to guard against response set bias. Table 1 identifies the construct name, the number of questions believed to load onto that construct, and a sample statement from each construct.

Table 1: The Eleven Theoretical Constructs of the OUS-I

Construct	*n*	Label	Sample Statement
Technology Usage	6	D	"I'd rather send a text than make a phone call."
Interface Preferences	1		"I don't think I would do well in Statistics if my class notes were on PowerPoint slides."
Math Skills	6	A	"I am comfortable doing square roots."
Software Skills	1		"I am willing to learn how to use any software that'll make my studies easier."
Time Management	3	H	"I learn the most the day before a test."

Study Skills	1		"I forget most of what I've learned right after the test."
Self-Starter / Isolationist	6	F	"I prefer to work alone when completing my classes."
Connectedness	3	E	"I will depend on study groups to get me through my online classes."
Way of Thinking	5	B	"I am completely comfortable with 'it depends' as my final answer.
Statistics Attitudes / Career Aspirations	5	C	"Knowing statistics will make me more marketable."
Reaching Out to Teacher	3	G	"I make appointments to talk with my instructor about my performance before I fall too far behind."

1.3.2 Data Collected

Table 1 shows that eight of the 11 theoretical constructs contained at least two statements. Despite the fact that three constructs had only one statement load onto it, data was collected for all 40 statements. However, the statements that loaded onto the *Interface Preferences, Software Skills* and *Study Skills* constructs were removed from any future analysis. The remaining 37 statements from the remaining eight constructs were analyzed and together make up the first iteration of the Online Undergraduate Statistics Inventory (OUS-I).

2. Psychometric Properties of the OUS-I

Since the OUS-I is being developed to predict success in online statistics courses, it is imperative that the inventory be multidimensional. At least two super constructs should be evident in any factor analysis of the inventory: one statistics construct and one online strategies construct. To this end, some statements were designed to specifically investigate the respondent's comfort, anxiety level, and understanding of the statistics subject. Still other questions were designed to test the participant's comfort operating in an online environment. As such, the OUS-I was hypothesized to be a multidimensional instrument. To confirm its multidimensionality structure, several dimensionality checks were performed.

Since the OUS-I included negative statements, the effect of the negative scoring was investigated. This is necessary to do since disagreeing with a negative statement is not the same as agreeing with a positive one. For example, if one disagrees with a statement that "all people are inherently bad," this does not mean that the same person would agree with a statement that "all people are inherently good."

Finally, since Likert data is not scale data, the effectiveness of the rating scale used must be established before the ordinal data can be converted to scale data. In the next three sections, the dimensionality, reverse coding, and rating scale effectiveness are discussed.

2.1 Dimensionality

To test the dimensionality of the 37-question OUS-I, six models were created. The decision about the OUS-I's dimensionality was guided by a comparison of each model's properties.

2.1.1 Unidimensional Model

Model 1 is a unidimensional model in which each of the eight constructs are independent manifestations of one super construct, generally referred to as "statistics strategies."

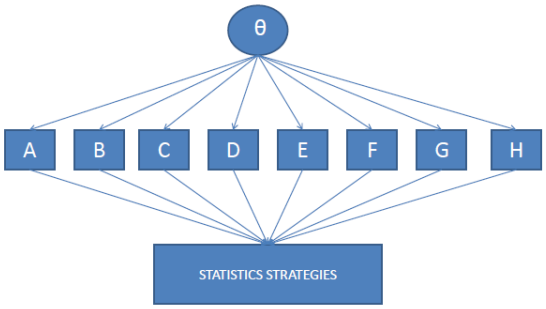

Figure 1: A schematic of the unidimensional model. Each letter represents the theoretical construct identified in Table 1.

2.1.2 Fully Saturated Eight-dimensional Model

Model 2 is an eight-dimensional model where each construct is thought to be a unique but correlated dimension.

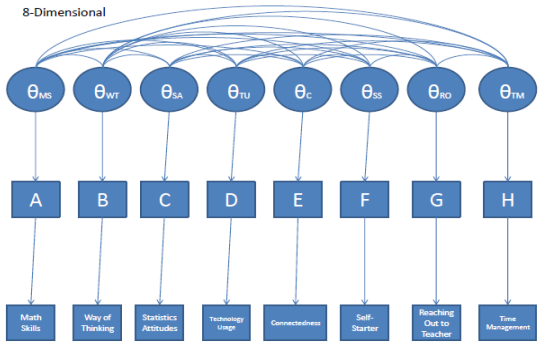

Figure 2: A schematic of the eight-dimensional model. Each letter represents the theoretical construct identified in Table 1.

2.1.3 Theoretical Two-dimensional Model

Model 3 is a theoretical 2D model in which one dimension is statistics-related and the other is related to online strategies. The two super-dimensions may be correlated.

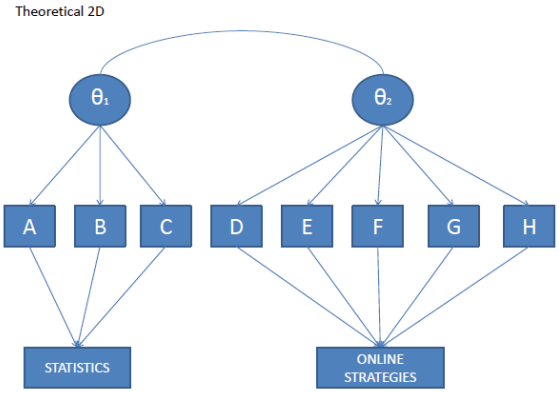

Figure 3: A schematic of the theoretical two-dimensional model. Each letter represents the theoretical construct identified in Table 1.

2.1.4 Theoretical Three-dimensional Model

Model 4 hypothesizes a three-dimensional model that isolates a statistics, online strategies, and general skills dimension. The three super-dimensions may be correlated.

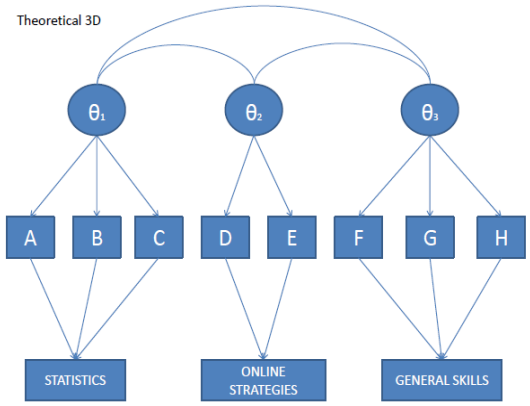

Figure 4: A schematic of the theoretical three-dimensional model. Each letter represents the theoretical construct identified in Table 1.

2.1.5 Correlations between the Eight Theoretical Constructs

Table 2 contains the correlations between the eight theoretical constructs. Constructs A, C, E, F, and D are highly correlated with each other and therefore form one empirical dimension while constructs B, G, and H form the other empirical dimension.

Table 2: Correlations between the Eight Theoretical Constructs

Correlations	A	B	C	D	E	F	G	H
A								
B	0.060							
C	0.198	0.052						
D	0.727	0.474	0.147					
E	0.809	0.121	0.536	0.836				
F	0.558	0.023	0.615	0.174	0.530			
G	0.067	0.890	0.192	0.427	0.219	0.293		
H	0.132	0.892	0.079	0.580	0.300	0.173	0.952	

2.1.6 Empirical Two-dimensional Model

Model 5 captures the two empirical dimensions identified in Table 2. At this time, neither of these empirical dimensions has been named.

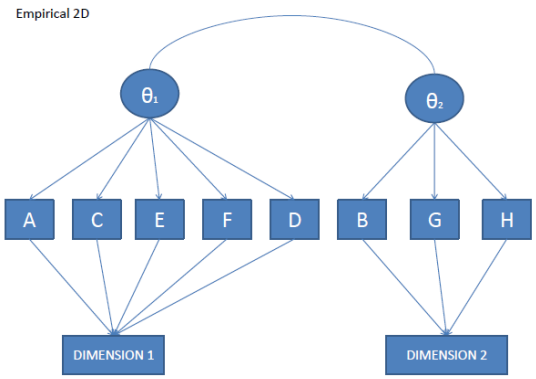

Figure 5: A schematic of the empirical two-dimensional model. Each letter represents the theoretical construct identified in Table 1.

2.1.7 Empirical 3-dimensional Model

Upon closer inspection of Table 2, it becomes clear that construct D (Technology Usage) loads highly onto both empirical dimensions (yellow and green dimensions). Model 6 therefore is the 3D empirical model with Technology Usage loading into one dimension.

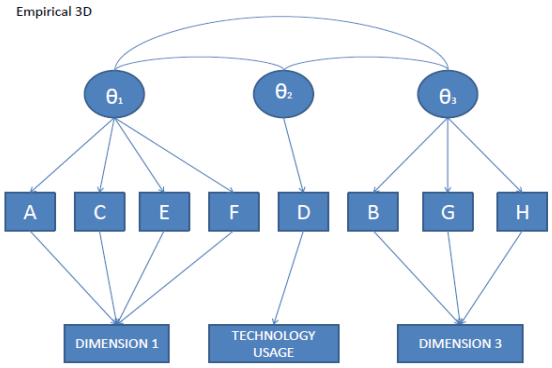

Figure 6: A schematic of the empirical three-dimensional model. Each letter represents the theoretical construct identified in Table 1.

2.1.8 Results

The data fit the multidimensional models better than the unidimensional model (see Table 3). The item reliability of the Theoretical 2D (Statistics and Online constructs) is the largest. On the other hand, the Empirical 2D model has a lower deviance than the Theoretical 2D model and has improved item reliability over the saturated Theoretical 8D model. These findings confirm that the OUS-I is a multidimensional instrument.

Table 3: Results from the Dimensionality Analysis of the Six Models

	Unidimensional	Theoretical (8D)	Empirical (3D)	Empirical (2D)	Theoretical (3D)	Theoretical (2D)
Deviance	13260.188	13032.477	13147.83	13167.71	13230.57	13260.32
Estimated Parameters	43	78	48	45	48	45
Item Reliability	0.986	0.984	0.984	0.986	0.985	0.987
Person Reliability	0.737	0.667 0.018 0.361 0.158 0.713 0.624 0.446 0.316	0.656 0.754 0.652	0.718 0.710	0.592 0.647 0.435	0.587 0.660
Number with Large Positive Outfit	4 of 37	6 of 37	5 of 37	5 of 37	4 of 37	3 of 37

2.2 Reverse Coding

To test the effectiveness of the OUS-I's reverse coding strategy, the statistics dimension from the Theoretical 2D model was analyzed. Table 4 provides a comparison between three models used to test the effectiveness of reverse coding. The first model is the current 16 question unidimensional model in which 11 statements are positive and 5

statements are negative. It is identified as "Unidimensional Reversed Codes" in the table. The second model is a multidimensional model in which the 11 positive statements load onto one dimension and the 5 negative statements load onto another. It is identified as "Multidimensional Positive and Negative Dimensions" on the table. The third model contains only the positive statements. It is identified as "Positive Statements."

2.2.1 Statistics Dimension

A two-dimensional model fit the 16 statistics items better than a unidimensional model that reverse coded negative items (smaller AIC and deviance). To this end, a model that fit only the 11 positive statistics statements fit the model much better than either of the other two models. This suggests that the negative items are tapping a different construct than the positive items. Adding a second correlated dimension does not improve model fit. As a result, the positive and negative items should not be combined into one scale. For the second iteration of the OUS-I, the negative statements will either be recoded positively or deleted altogether.

Table 4: Results from the Effectiveness of the Reverse Coding Used in the Theoretical Statistics Dimension

Statistics Dimension	Unidimensional Reversed Codes	Multidimensional Positive and Negative Dimensions	Positive Statements
Schematic	++++++++++ -----	++++++++++ -----	++++++++++
Deviance	5614.80299	5500.10810	4076.53624
Number of Estimated Parameters	22	28	18
Item Separation	0.973	0.938	0.964
AIC	5658.80299	5548.10811	
Change in Deviance Change in Degrees of Freedom	114.69489 6		1423.57186 10
Critical Value (α = .01)	16.81		23.21
Interpretation	Adding the second correlated dimension improves the model fit. As a result, the positive and negative items should NOT be combined into one scale.		
Decision	The sum of the 11 positive statistics statements will be considered the measure of the statistics dimension.		

2.2.2 Online Strategies Dimension

A similar analysis was performed for the online strategies dimension (Table 5). This dimension contained 18 positive statements and 3 negative statements. Again, the negative Online Strategies items appear to be tapping a different construct than the

positive items. This is evident by the lower deviance value for the multidimensional model as compared to the unidimensional model.

Table 5: Results from the Effectiveness of the Reverse Coding Used in the Theoretical Online Strategies Dimension

Online Strategies Dimension	Unidimensional Reversed Codes	Multidimensional Positive and Negative Dimensions
Schematic	++++++++++ ++++++++---	++++++++++ ++++++++ ---
Deviance	7485.78947	7385.47784
Number of Estimated Parameters	27	31
Item Separation	0.977	1.000
Change in Deviance Change in Degrees of Freedom	100.31163 4	
Critical Value (α = .01)	13.28	
Interpretation	Adding the second correlated dimension improves the model fit. As a result, the positive and negative items should NOT be combined into one scale.	
Decision	The sum of the 18 positive statistics statements will be considered the measure of the online strategies dimension.	

2.3 Rating Scale

The OUS-I captures ordinal Likert data. This ordinal data must be transformed into a quantitative measure before any data analysis requiring quantitative data is performed. Verifying the effectiveness of the rating scale is an important step into converting ordinal response into scale measures. A well-behaved rating scale has at least ten responses in each category, ordered measures and thresholds, and mean outfit statistics of less than two. Table 6 provides the results of the OUS-I's rating scale analysis.

Table 6: Results of the OUS-I Rating Scale Analysis

Original OUS-I Rating Scale					
Category	Label	Observed Count	Average Respondent Measure	Outfit Mean Square Fit Statistic	Threshold
1	Strongly Disagree	32	-.44	1.50	NONE
2	Disagree	78	-.08	1.62	-1.31
3	Somewhat Disagree	84	-.13	.72	-.26
4	Neither Agree nor Disagree	169	.13	1.01	-.63
5	Somewhat Agree	299	.42	.99	-.21
6	Agree	320	.94	.76	.62
7	Strongly Agree	164	1.46	.97	1.79

The original seven point scale (using data from the statistics dimension) has sufficient counts as each category contains at least ten responses. The average measures are also mostly well-behaved; however, the third category's measure is slightly smaller than the second category's measure. Hence, the average measures are disordered. The mean square fit statistics are all less than two in absolute value. Finally, the thresholds based on this seven point scale are disordered. Indeed, the fourth threshold is significantly smaller than the third threshold.

This disordering is evident in a graph of category probabilities versus measure relative to item difficulty for the OUS-I statement, "I am completely comfortable with 'It Depends' as my final answer" (Figure 7). Notice how the magenta graph is never the highest peak at any item difficulty level.

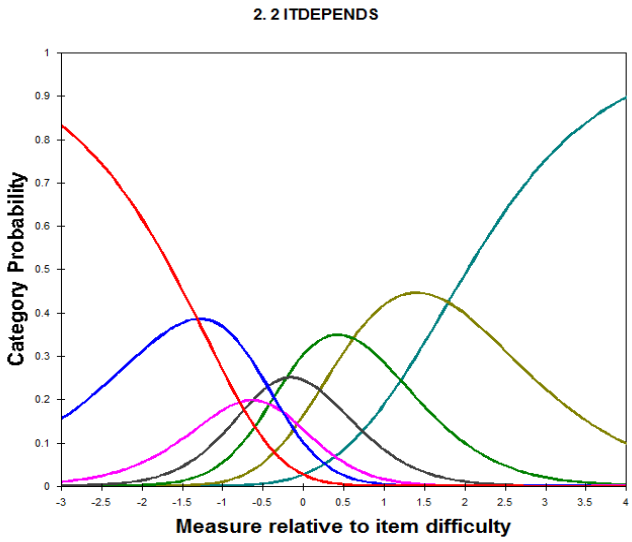

Figure 7: Original ITDEPENDS Statement: Threshold Analysis

In an effort to order the thresholds of the rating scale, categories were recoded, switched, and collapsed. Seven alternate rating scales were hypothesized and tested (Table 7).

Table 7: Investigating the Thresholds of Alternate Rating Scales

Rating Scale	Scale Size	Ordered Thresholds?
1 2 3 3 3 4 5	5	NO
1 1 2 3 4 5 5	5	NO
1 1 1 2 3 3 3	3	NO
1 1 2 2 2 3 3	3	YES
1 2 2 3 4 4 5	5	NO
1 2 4 3 5 6 7	7	NO
1 2 3 3 4 5 6	6	YES

Thresholds became ordered in a 3-point and 6-point scale. The 3-point scale combined the lowest two, middle three, and highest two categories. The original 7-point scale can therefore be recoded to create a 3-point scale (Disagree-Neutral-Agree). Figure 8 shows the thresholds of the "It Depends" statement using a 3-point scale. Notice how each category has an opportunity to have the highest probability of selection.

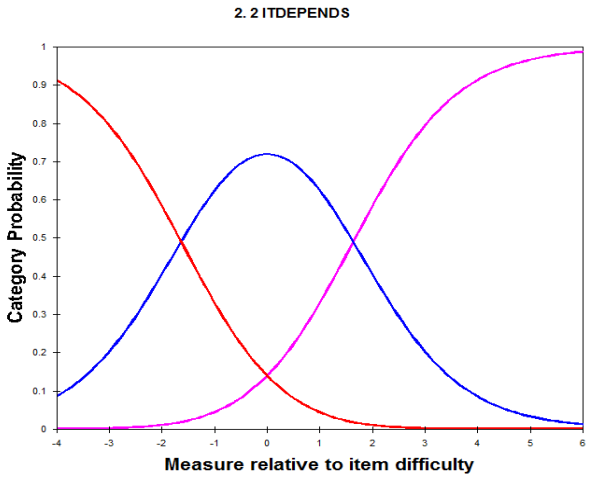

Figure 8: 3-Point ITDEPENDS Statement: Threshold Analysis

The 6-point scale combined the somewhat disagree and neutral categories. Using this scale, the middle neutral category is combined with the "Somewhat Disagree" category. Figure 9 shows the ordered nature of the 6-point solution. In this rating scale, similar to the 3-point scale, each category has an opportunity to have the highest probability of being selected.

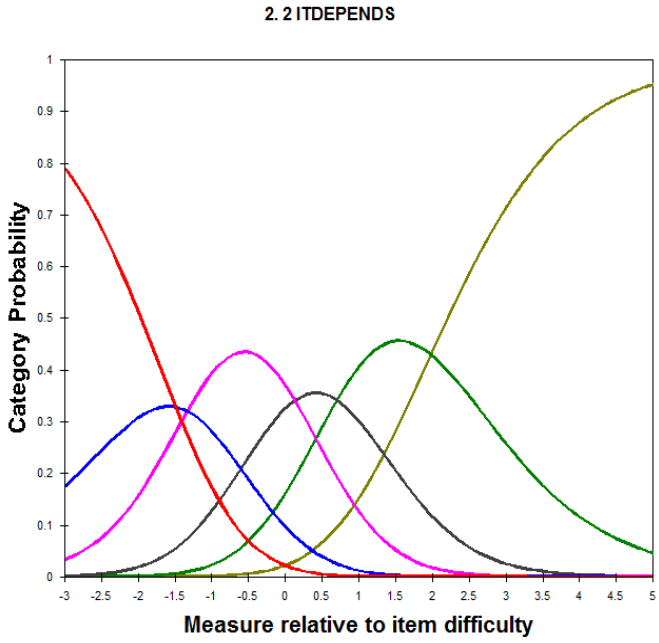

Figure 9: 6-Point ITDEPENDS Statement: Threshold Analysis

3. Next Steps

There is still much to be done before the OUS-I is ready to be released to the public. First, the effectiveness of the rating scale must be determined for the "Online Strategies" data. Secondly, negatively phrased statements will either be deleted or reversed to become positive statements in the next OUS-I iteration. In addition, to determine which rating scale is optimal, two new versions of the OUS-I will be developed using a 6-point scale and a 3-point scale. Additional data will be collected over the next three semesters. Using the results from the second version of the OUS-I, the following questions will hopefully be addressed:

- Does the data fit the 3-point or 6-point scale better?
- Should negatively phrased items be included in the scales?
- Does the data better fit the Empirical 2D or Theoretical 2D model?
- Can a theory be developed to link the items in the Empirical 2D model?

After the psychometric properties of the second version of the instrument have been fully determined, scores on the statistics and online strategies dimensions and the original eight theoretical constructs will be used to predict academic performance in the fully online statistics course.

Acknowledgements

The author would like to thank Dr. Rick Simmons (Texas A&M University Central Texas), Instructor Megan Mocko (University of Florida), and their students along with his own students (Shorter University, Georgia State University) who provided the data used to study the psychometric properties of the OUS_I.

References

Smith, Jr., E.V. & Smith, R. M. (Eds.). (2004). *Introduction to Rasch Measurement.* Maple Grove, MN: JAM Press.

Smith, Jr., E.V. & Smith, R. M. (Eds.). (2007). *Rasch Measurement: Advanced and Specialized Applications.* Maple Grove, MN: JAM Press.