Psychometric Analysis (August 2017)

Reliability and Validity of Sheridan's New Course Evaluation Survey Tool

Notes:

- *Some content used in this document is adapted from the University of Toronto's Course evaluation guidebook and resource manual (Rolheiser, C., Werhun, C. D., & Gravestock, P. (2017).*
- *All data, analyses and test results are based on Sheridan College (FAST only) course evaluation data collected from Fall 2016 to Winter 2017.*
- *Sheridan's 5 core course evaluation items are referenced throughout this document. Please refer to the table below for item wording.*

1	Overall, the quality of my learning experience in this course was…
2	The course provided an appropriate balance between instruction and practice
3	The professor respected different opinions and experiences
4	The course helped me develop aspects of my creative abilities (e.g., thinking, writing, inventing, designing, performing, etc.)
5	SLATE was effectively used to organize this course

2017 Factor Analyses

Faculties: FAST

N = 14,769 evaluations

Core Course Evaluation Items:

Factor Analysis

This analysis is of particular importance within a context like Sheridan's, who wishes to calculate a composite (or overall grand mean) of the first five items. In essence, factor analyses gauge the extent to which individual items can be meaningfully summarized by a single larger category.

To ensure factor analyses were appropriate within the sample, a preliminary test of **sampling adequacy**, the KMO (Kaiser-Meyer-Olkin) test, which calculates the ratio of the squared correlation between items to the squared partial correlation between variables, generated a "marvelous" outcome of 0.9. This value and associated rating suggests that further factor analyses should generate "distinct and reliable factors" based on relationships between the items, justifying the use of this analysis (Kaiser, 1974).

Factor analyses with varimax rotation indicated that core course evaluation items are explained by a single underlying factor, accounting for 77% variance (Eigenvalue = 3.8); each item loading ranges from .8 to .9 with single factor.

Item	Factor
1	.91
2	.91
3	.89
4	.87
5	.79

Tests of Internal Consistency

Faculties: FAST

N = 14,769 evaluations

Core Course Evaluation Items:

Tests of Internal Consistency

- Tests of internal consistency and reliability reveal strong interrelatedness among the core course evaluation items, with an overall Cronbach's alpha = .97, which is highly acceptable given the number of items (*n* = 5; Cortina, 1993).
- Examination of the impact of items deleted on Cronbach's alpha reveal no meaningful gains in the statistic when any items are removed. Also, corrected item-total correlations reveal that each item correlates similarly and strongly with the total scale. Specifically, each core item correlates >.7 with the overall scale, suggesting further that each of the core items should be retained within the scale (Field, 2013).
- The analysis does suggest somewhat of a unique relationship between item 5 (the SLATE item) and the overall scale, as this item relates less strongly than the other core items do (as noted in the corrected item-total correlations).
    - Item 5's content draws student attention to the use of technology – SLATE – as an organizational tool within the course. In comparison to the content of the other four core items, it is reasonable that the psychological factors that come into students' survey responses to this item are slightly unique to those that emerge in students' responses to the other core items. Specifically, whereas item 5 is indeed related strongly to the overall scale, suggesting the inherent interrelatedness between general course and general instructor items, it is reasonable that students think somewhat uniquely about the role of technology in the course (compared to the other course components), but not in a manner that demonstrates statistical concern.

- To rule out any concern that item 5 is appreciably unique from the other core items, a simple regression analysis, regressing items 1, 2, 3, and 4 on item 5 was conducted, which revealed strong correlations between all core items and item 5 (*r* range .61 to .63), (R = .69, R^2 = .48, F (4, 14,356) = 3,343, p < .01). This further suggests that students' somewhat disparate consideration of the content of item 5 is not meaningfully different to their consideration of the content of the other core course evaluation items. Please refer to the tables below for the regression coefficients.

b. *Tests of Internal Consistency of Core Items*

Item	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	.85	.89
2	.86	.89
3	.79	.91
4	.83	.90
5	.69	.93

c. *Regression of Core Items 1, 2, 3, and 4 on Item 5*

Item	*B*	*t*	*r* (zero-order with item 5)
1	.10	9.85	.63
2	.28	19.78	.64
3	.21	20.79	.61
4	.21	20.51	.63

All *t* values significant at p <.001.

Relationship to Overall Learning Experience

Faculties: FAST

N = 14,769 evaluations

Core Course Evaluation Items:

Relationship to Overall Learning Experience

To further test how the core course evaluation items relate to students' perceptions of their overall learning experiences within their courses, a multiple regression was conducted on the core items (2 to 5) as predictors of item 1, students' rating of their overall learning experience within the course (from Poor to Excellent).

Analyses revealed that the variance in students' perceptions of their overall learning experiences is explained significantly by the five item predictors of the model, R^2 = .75, F (4, 14,353) = 10,585.77, p < .01. Moreover, no concerns regarding multicollinearity among the predictors emerged, as all estimates of Tolerance were >.3 and estimates of Variance Inflation Factor (VIF) were <4, which are within acceptable ranges.

With respect to the influence of each individual item on students' overall learning experiences within their courses, please refer to the table below. Here, analyses reveal that each item predicts students' perceptions of their overall learning experiences, as indicated by the significance of each item beta weight; recognizing some relationships vary in estimate strength with the criterion, overall learning experience. Specifically, analyses reveal a unique relationship between item 5 (i.e. the extent to which SLATE was used to organize the course) and the criterion, item 1, as indicated by the partial correlation of .08 and the part correlation of .04 in the table below.

The data also suggest that students' perception of the extent to which the course provides balance between learning and practice plays an important, and somewhat more influential, role in their ratings of their overall learning experience in their courses.

d. *Regression of Core Items on Item 1 (Rating of Overall Learning Experience)*

			Correlations		
			zero-order	partial	part
2	.48	55.71	.82	.42	.23
3	.23	27.71	.74	.23	.11
4	.33	40.57	.79	.32	.17
5	.07	9.86	.63	.08	.04

All t values significant at p < .01.

Convergent & Concurrent Validity

Concurrent Validity

Faculties: FAST

N = 14,769 evaluations

CORE ITEM AND ITEM BANK RELATIONSHIPS: CONCURRENT VALIDITY

To substantiate further the validity of the core and Item Bank items, tests of concurrent validity were conducted. Specifically, we investigated the extent to which scores on one item would relate meaningfully differently on other items. To do this, a tripartite split was applied to Core Item 1 to establish 3 groups of scores on perceived quality of learning experience: High, Mid, and Low (High = mean scores greater than 4.5; Mid = mean scores less than or equal to 4.5 and greater than or equal to 3.5; Low = mean scores less than 3.5). This grouping of mean scores on Core Item 1 was then used to compare responses on Item Bank items. analyses confirm that perceptions of overall quality of learning experiences within a course are associated empirically with expected theoretical differences in mean scores on Item Bank items. In other words, low scores on perceived quality of overall learning experience are associated, as one would predict, with low scores on Item Bank Items; high scores on perceived quality of overall learning experience are associated with highest scores on Item Bank items; and mid scores on perceived quality of overall learning experience are associated with mid-range scores on Item Bank items. Please refer to the table below.

		Mean Category		
		H	**M**	**L**
Item bank Item #				
Q19		4.7	4.4	3.7
Q20		4.8	4.3	3.4
Q21		4.6	4.3	4.5
Q22		4.5	4.5	4.6
Q23		4.8	4.4	3.9
Q24		4.8	4.5	3.9
Q25		4.8	4.3	3.7
Q26		4.8	4.4	3.5
Q27		4.7	4.7	
Q28		5.0	4.4	3.4
Q29		4.6	4.0	
Q30		4.8	4.3	3.5
Q31		4.8	4.3	3.2

Q32				3.8
Q33		4.7	4.0	3.7
Q34		4.8	4.4	3.9
Q35		4.7	4.3	3.4
Q36		4.9	4.2	3.3
Q37		4.3	4.0	4.1
Q38				
Q39		4.6	4.3	3.3
Q40		5.0	4.2	
Q41		5.0	4.3	
Q42		5.0	4.3	2.1
Q43		4.8	4.3	3.4
Q44		4.5	4.3	2.9
Q45				
Q46		4.7	4.2	3.2
Q47			4.1	
Q48		4.5	4.2	
Q49		4.7	4.5	3.0
Q52		4.4	4.4	4.0
Q53		4.8	4.2	3.7
Q54		4.5	4.1	4.0
Q55		5.0	3.9	3.8
Q56		4.7	4.4	3.5
Q57		5.0	4.3	3.5
Q59		4.5	4.3	3.4
Q60		4.9	4.4	3.1
Q61		4.8	4.8	
Q62			3.9	
Q64		4.8	4.0	
Q65		5.0		4.0
Q81		4.6	3.9	4.3
Q108				3.0
Q109				3.0
Q110				
Q111				3.0
Q112				
Q113				
Q114				
Q115		5.0		3.9
Q116			4.4	
Q117				
Q118				
Q119		4.7	3.7	
Q120				3.3
Q141		4.8	4.3	

Q142			3.7	2.9
Q143				3.9
Q144				3.1
Q145				
Q146				2.5
Q147				2.5
Q148		4.7	4.2	
Q149		4.9	4.5	3.2
Q150				
Q151				
Q152		4.5		
Q153			4.3	
Q154				
Q155				
Q156			3.8	2.9
Q157				
Q158			4.2	
Q159				
Q160			4.2	
Q177				4.1
Q178				
Q179				3.0
Q180				
Q181				
Q182				
Q183				
Q184				
Q185				
Q186				
Q187				
Q189				
Q190				
Q191				
Q192				
Q193				
Q194				
Q195				
Q196		3.3	3.2	
Q197		5.0		
Q198				
Q199		4.5		
Q200		4.6		4.0
Q201		4.5		
Q202		4.8	4.4	
Q203		4.7	3.7	

Q204 Q205		4.9	5.0	
Averages		**4.7**	**4.2**	**3.5**

Convergent Validity – *Analysis to be completed once sufficient data have been collected*

Faculties: FAST

N = 14,769 evaluations

RELATIONSHIPS BETWEEN CORE AND ITEM BANK ITEMS

Establishing convergent validity of evaluation items provides additional assurance that the core course evaluation items are measuring what they were intended to measure. By definition, convergent validity reflects the extent to which different survey items that measure distinct, but theoretically related concepts, are correlated or related, which is an important requisite for item validation. To illustrate, for example, when validating survey items intended to gauge the extent to which people are afraid of heights, we would expect that other items that measure the extent to which people are fond of views from high places or the extent to which they like sitting in a window seat on an airplane, would be related meaningfully to our “fear of heights” items. Thus, uncovering theoretically and pedagogically meaningful relationships between the core and Item Bank items further substantiates their relevance within Sheridan’s overall assessment framework.

To that end, this next section summarizes convergent relationships between the core course evaluation items and Item Bank items. Given that Item Bank items vary in selection and usage by instructors, which thus determines the amount of responses collected, we report convergent relationships if they meet threshold criteria of sample size (n > 900 responses) and relationship strength (| r | > .4).