SLS 1101 Course Level Assessment Report – Spring 2017

Author: Joseph F. van Gaalen, Ph.D., Director, Academic Affairs Assessment

Introduction

Florida SouthWestern State College's Academic Success Department assessment plan currently includes collection of achievement data to provide a baseline measure of the efficacy of the SLS 1101 *College Success Skills*. The assessment outcomes are intended to provide a baseline and measurement of achievement moving forward as well as investigate the strength and performance of items in the exam. The assessment plan also provides comparisons between dual Enrollment and non-dual enrollment students, online versus traditional students, and by site, where possible. Where data is sufficient, additional analyses are provided including distribution studies and longitudinal studies.

For additional detail or further analysis not provided in this report, please contact Dr. Joseph F. van Gaalen, Director of Academic Assessment, Academic Affairs (jfvangaalen@fsw.edu; x16965).

Assessment Item Analysis

Insight into the strength of the assessment questions offers information on student learning and helps to discriminate those students which have learned the material and those that did not (Ding and Beichner, 2009). Item analysis measures the difficulty of the question, attempts to define the capacity of the question to discriminate between higher achieving from lower achieving students, and the reliability of the questions for measuring common materials (Doran, 1980).

The SLS 1101 common course assessment consists of 28 multiple choice questions and an additional set of multiple response questions. Item difficulty for each of the multiple choice questions (i.e., the first 28 questions only) was calculated using standard practices whereby the higher the value the easier the question (Ding and Beichner, 2009). A score of 1 means responses from all test takers were correct responses while a score of 0 means none were correct responses. A score of 0.5 ought to be a goal for the assessment author such that the question is neither too easy, nor too hard, so as to effectively discriminate between students of robust knowledge and those lacking (Ding and Beichner, 2009; Doran, 1980). A more detailed interpretation of item difficulty is presented in Table 1.

Difficulty	Range of Values
Very easy	0.85 – 1
Moderately easy	0.60 – 0.85
Moderately difficult	0.35 – 0.60
Very difficult	0.00 – 0.35

Table 1. Item difficulty interpretation as defined by Doran (1980).

Figure 1 depicts item difficulty results for the SLS 1101 common assessment. A total of 19 of 28 questions exhibit poor item difficult scores. Questions 2-4, 7-9, 11, 15-23, 25, 26, and 28 exhibit item difficulty categorized as 'too easy' according to literature (Doran, 1980). Question difficulty should match the intent of the test in conjunction with corresponding levels of both student and course level (Doran, 1980). Adjustments should be made with this in mind.

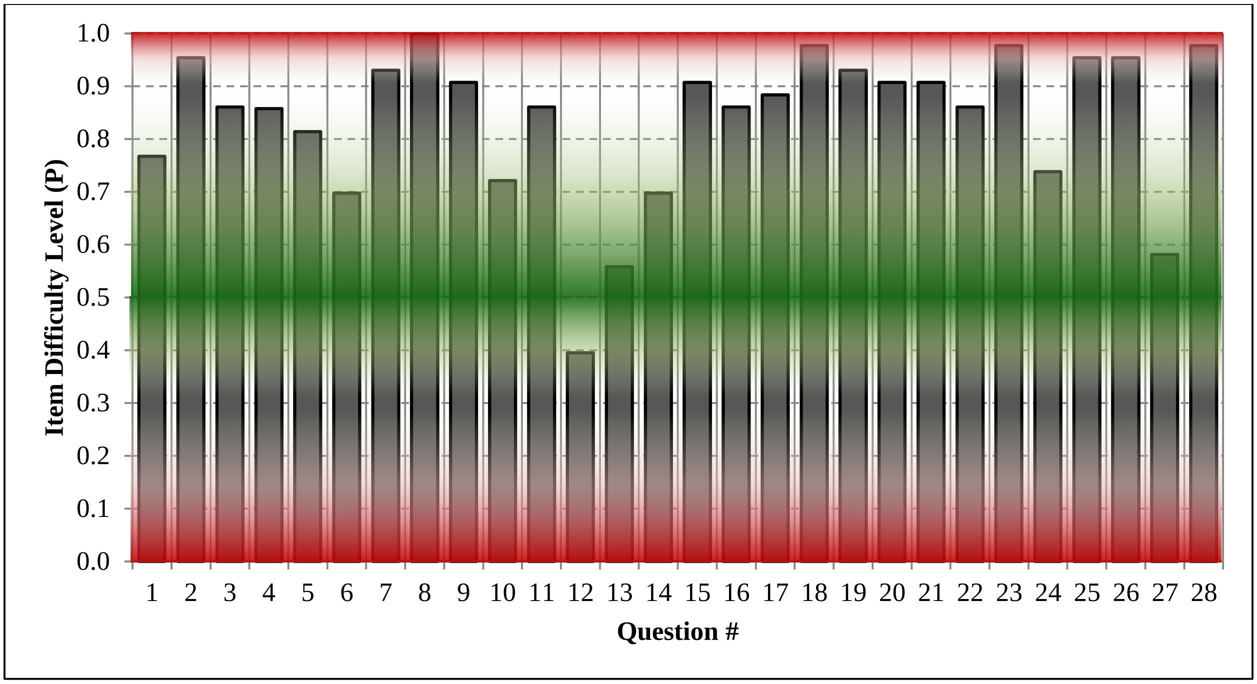

Figure 1. Item difficulty results for SLS 1101 common final. Green shaded regions depict moderate questions with darker green centered on the generally accepted 0.5. Lighter green fading to red depict questions progressively more difficult (near 0) or easier (near 1).

An item discrimination index was calculated for each of the assessment questions. The item discrimination index measures performance of an item with respect to the most successful (upper quartile) and least successful (lower quartile) of the class. Results can be used to determine how well an item discriminates between high performing students and low performing students. This technique can be done using external quartiles (e.g. student overall course grade or some other performance representation) or internal (the overall score on the test in question). In this case, an internal criterion was used.

A generally accepted cutoff for valuing a question as adequately discriminating is 0.3, however, this cutoff is arbitrary and so whether a question is suitably discriminating or not is at the discretion of the test author (Doran, 1980). Table 2 presents a guide for discriminating question indices. Note that a negative discrimination index would mean the lower quartile of students is more likely to correctly answer the questions and so these, if they occur, should be closely examined immediately.

Discrimination	**Range of Values**
Very strongly discriminating	> 0.6
Strongly discriminating	0.4 – 0.6
Moderately discriminating	0.2 – 0.4
Weakly discriminating	0.1 – 0.2
Very weakly discriminating	0 – 0.1

Table 2. Item discrimination index interpretation (Doran, 1980).

Figure 2 depicts item discrimination index results for the SLS 1101 common course assessment. The graph has a gradient such that questions calculated as strong discriminators (see Table 2) fall in the green shaded region fading to white for moderate discriminators, and eventually to red for weak discriminators. A total of 17 of 28 questions exhibit poor item discrimination. Questions 1, 2, 4, 7, 8, 11, 15-19, 21-23, 25, 26, and 28 are is considered weakly discriminating according to literature (Doran, 1980).

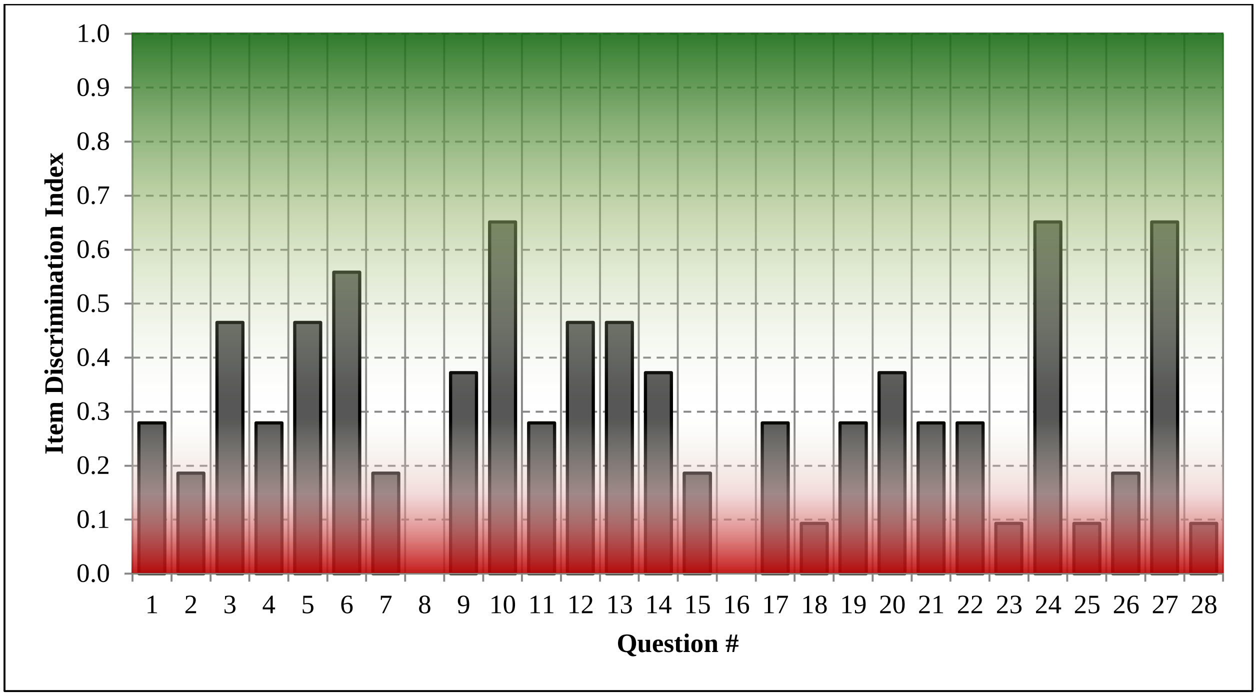

Figure 2. Item discrimination index results for SLS 1101 common assessment. Green shaded regions depict progressively stronger discriminators. White shaded regions depict moderately strong (>0.3) to moderately weak (<0.3) discriminators. Red shaded regions depict weak discriminators.

Item discrimination indices are a measure of the discrimination of the correct option with respect to other options available yielding no information specific to any one discriminator themselves. Therefore, when reviewing discrimination index results with an eye towards revision it is important to review the nature of the distractors, in particular those items which may be questionable (Ding and Beichner, 2009; Doran, 1980). Figure 3 and the associated Table 3 depict the percentage of response selection for each answer choice by question. A total of 1 of 28 questions exhibits a greater response rate for the 1st distractor than for the correct response. Question 12 exhibit greater response rates for the 1st distractor than for the correct response.

In addition to item difficulty and discrimination index, a Point Biserial Index (PBI) was calculated. The PBI measures the reliability of the item compared with score distribution (Ding and Beichner, 2009). In other words, a low PBI is indicative of an item either not testing the same material or not testing it in the same manner or level, since questions on the same test ought to be testing material within the same domain as the other questions on that test. An item with a PBI of greater than or equal to 0.2 indicates the item is performing similar to that of its counterparts. A PBI lower than 0.2 indicates the material is not strongly linked with other items and may require review to ensure the efficacy of the question (Ding and Beichner, 2009).

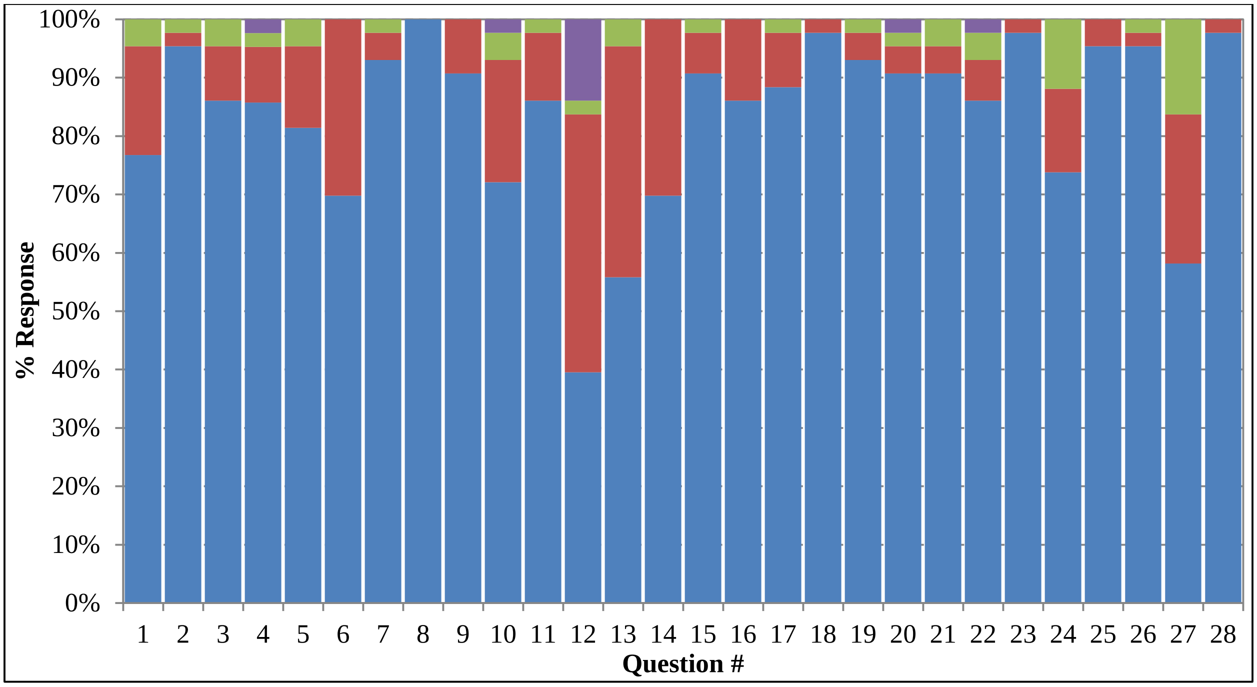

Figure 3. Item response distribution for SLS 1101 depicting selection percentages of each response option for each question. Blue denotes correct responses, red the most commonly selected distractor, green the 2nd most commonly selected distractor, purple the 3rd most commonly selected distractor, and if applicable, light blue the least commonly selected distractor.

Q	Correct	1st Distractor	2nd Distractor	3rd Distractor
1	A	C	D	B
2	A	D	C	B
3	A	C	D	B
4	B	A	C	D
5	C	B	A	D
6	D	A	B	C
7	D	B	C	A
8	C	A	B	D
9	D	A	B	C
10	B	D	C	A
11	B	A	D	C
12	C	D	B	A
13	A	D	C	B
14	A	B	C	D
15	C	D	B	A
16	D	B	A	C
17	B	D	C	A
18	A	C	D	B
19	B	A	D	C
20	B	A	D	C
21	D	A	B	C
22	D	B	C	A
23	B	C	A	D
24	C	D	B	A
25	A	B	C	D
26	C	B	D	A
27	C	A	B	D
28	A	C	B	D

Table 3. Response option for corresponding correct responses and distractors for SLS 1101. Cell colors reflect representation in Figure 3 above.

Figure 4 depicts the PBI results for the SLS 1101 common course assessment. The graph has a gradient such that questions calculated with strong reliability fall in the green shaded region fading to white for moderate to strong reliability, and eventually to red for weak reliability. A total of 5 of 28 questions exhibit low PBI. Questions 1, 2, 4, 8, and 13 exhibit a low PBI meaning they are considered potentially unreliable according to literature (Ding and Beichner, 2009). Note that variability in question length, vocabulary, clarity, strength of distractors, potentially interconnected questions (clues for one question found in another), and option logic all have the potential to cause variability in PBI, should topical correlations exist (Suskie, 2004). A more thorough appraisal of the many considerations of writing a multiple choice and true/false assessment can be found on pages 200-211 of the Suskie (2004) work.

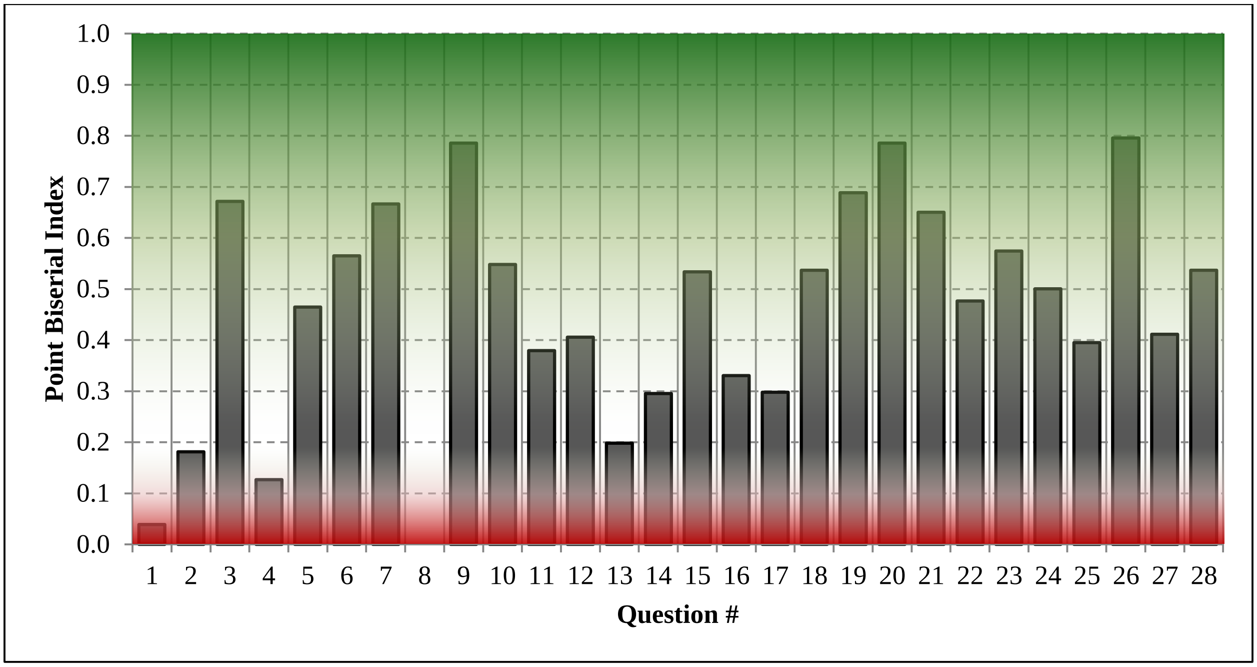

Figure 4. Point Biserial Index (PBI) results for SLS 1101. Green shaded regions depict progressively stronger reliability. White shaded regions depict moderately strong reliability. Red shaded regions depict weak reliability.

A full list of questions which scored poorly in each of the three item analyses is shown in Table 4. Questions 2, 4, and 8 exhibit poor item analytics in all three areas. Questions 1, 7, 11, 15-19, 21-23, 25, 26, and 28 exhibit poor item analytics in two of three areas. In total, 19/28 questions exhibit poor item difficulty scores, 17/28 exhibit weak item discrimination indices, and 5/28 exhibit low PBI for a total of 21/28 unique questions exhibiting poor scores in any one area.

Item	Item Difficulty	Item discrimination index	PBI
Q1		Weak	Low
Q2	Too easy	Weak	Low
Q3	Too easy		
Q4	Too easy	Weak	Low
Q7	Too easy	Weak	
Q8	Too easy	Weak	Low
Q9	Too easy		
Q11	Too easy	Weak	
Q13			Low
Q15	Too easy	Weak	
Q16	Too easy	Weak	
Q17	Too easy	Weak	
Q18	Too easy	Weak	
Q19	Too easy	Weak	
Q20	Too easy		
Q21	Too easy	Weak	
Q22	Too easy	Weak	
Q23	Too easy	Weak	
Q25	Too easy	Weak	
Q26	Too easy	Weak	
Q28	Too easy	Weak	

Table 4. List of items that are outside the generally accepted scores indicating a strong multiple choice question.

Descriptive Statistics and Learning Objectives

The FSW Academic Success Department established the common course assessment to include 28 multiple choice questions. The results are analyzed to assess student general knowledge and to make comparisons between ground, online, and dual enrollment students. The results will be used to establish baseline data for defining future Student Learning Objectives (SLOs) in moving forward.

For the spring 2017 assessment, 43 artifacts were collected for SLS 1101 from 2 of 3 course sections. The third section did not administer the assessment. Descriptive statistics for SLS 1101 artifacts are shown in Table 5. For a graphical representation of score distribution, see Figure 5. Artifacts scores are centered on 26/28 and are normal exhibiting a large negative skew, or slight shift of the peak towards higher scores.

Maximum score	*28*
n	43
Max	28
Min	8
Median	25
Mode	26
Mean	23.4
Standard deviation	4.14
Skewness	-2.34
Kurtosis	6.45

Table 5. Descriptive statistics for SLS 1101 common course assessment.

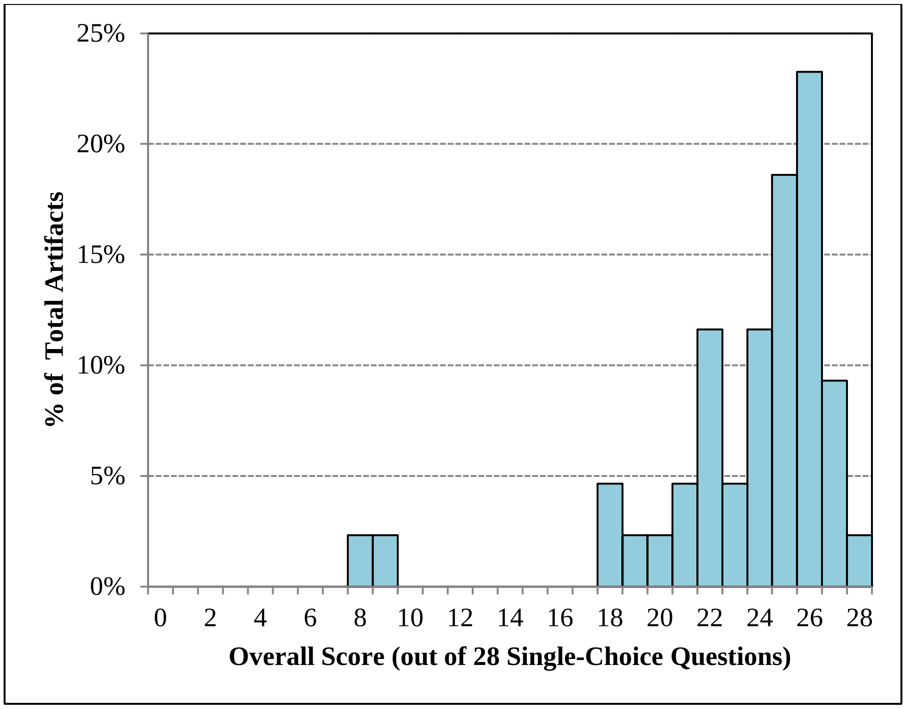

Figure 5. Score distribution for SLS 1101 artifacts.

Exploratory Analysis and Significance Testing

Multiple comparisons of artifact scores across varying formats, campuses, and student types were made, where possible, in order to add depth to the causes of the distribution of the artifacts. Each course was divided into the appropriate subgroups to perform the analysis. In cases where a subgroup is not represented in the course comparisons were not conducted and are noted for comprehensiveness.

Dual Enrollment to non-Dual Enrollment Comparison

All sections of the course offered during spring 2017 were offsite dual enrollment sections therefore no comparison of dual enrollment and non-dual enrollment could be made.

Online to Traditional Comparison

No online sections of the course were offered during spring 2017 therefore no comparison between online and traditional could be made.

Comparison by Campus/Site

All course sections during spring 2017 were offered offsite (dual enrollment). Therefore, no cross-campus comparison could be made.

Longitudinal Study

Item Analysis

Item analysis results vary with the performance on an assessment. As such, it is necessary to review results over time to discovery potential patterns or trends existing within the data. For brevity, detailed below in Table 6 are the common poor performing scores through time in item difficulty, discrimination index, and PBI. Questions 1, 13, 15, 16, 17, 20, and 21 exhibit poor performance over both terms over the course of the study.

Item	Fall 2016	Spring 2017
Q1	Too easy, Weak discriminator, Low PBI	Weak discriminator, Low PBI
Q2		Too easy, Weak discriminator, Low PBI
Q3		Too easy
Q4		Too easy, Weak discriminator, Low PBI
Q5	Too easy, Weak discriminator, Low PBI	
Q7		Too easy, Weak discriminator
Q8		Too easy, Weak discriminator, Low PBI
Q9		Too easy
Q11		Too easy, Weak discriminator
Q13	Weak discriminator, Low PBI	Low PBI
Q14	Weak discriminator, Low PBI	
Q15	Too easy, Weak discriminator, Low PBI	Too easy, Weak discriminator
Q16	Too easy, Weak discriminator, Low PBI	Too easy, Weak discriminator
Q17	Too easy, Weak discriminator, Low PBI	Too easy, Weak discriminator
Q18		Too easy, Weak discriminator
Q19		Too easy, Weak discriminator
Q20	Too easy, Weak discriminator, Low PBI	Too easy
Q21	Too easy, Weak discriminator, Low PBI	Too easy, Weak discriminator
Q22		Too easy, Weak discriminator
Q23		Too easy, Weak discriminator
Q25		Too easy, Weak discriminator
Q26		Too easy, Weak discriminator
Q28		Too easy, Weak discriminator

Table 6. List of items that are outside the generally accepted scores of a strong multiple choice question for SLS 1101 for fall 2016 through spring 2017.

Data Distribution

Description of achievement over time in SLS 1101 is provided in Figure 6. Each term exhibits similar distribution characteristics. Both are very negatively skewed and both exhibit a small distribution at the lower end of the scoring range. Note that comparison from fall-to-spring may be less useful as assessment reports across multiple course level and program level assessments at FSW typically exhibit substantial differences from fall to spring term and are better interpreted from fall-to-fall and spring-to-spring (see http://www.fsw.edu/facultystaff/assessment/history for further details).

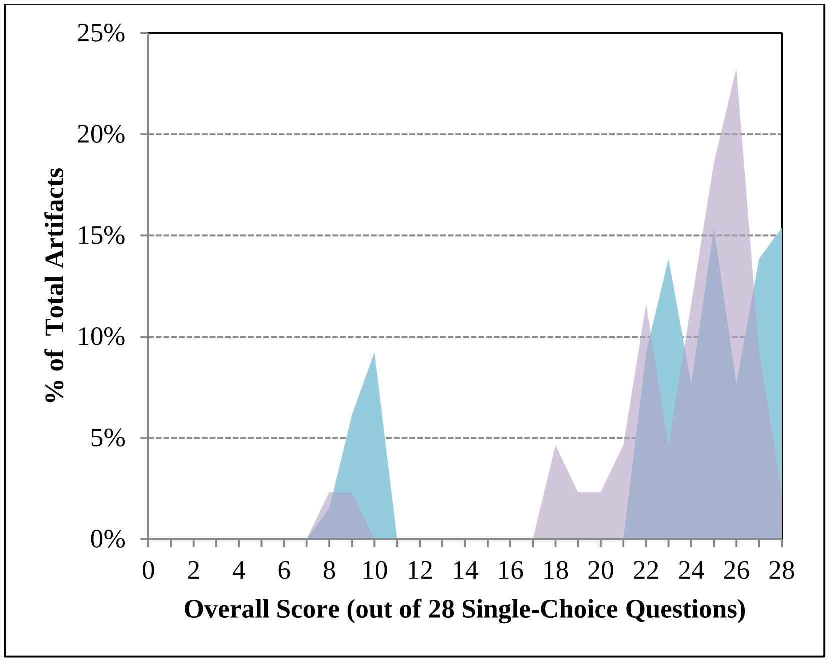

Figure 6. Comparison of score distributions for fall 2016 (teal) and spring 2017 (purple).

Conclusions

FSW's Academic Success Department has employed a common course assessment to assess SLS 1101 *College Success Skills*. This report provides analysis of the results as well as the effectiveness of the assessment in measuring learning in an effort to continuously improve assessment strength.

A drill-down of SLS 1101 results are as follows:

1. In an item analysis of the 11 questions in the common course assessment a total of 19 of 28 questions exhibit poor item difficult scores. Questions 2-4, 7-9, 11, 15-23, 25, 26, and 28 exhibit item difficulty categorized as 'too easy' according to standards.
2. In the same item analysis, a total of 17 of 28 questions exhibit poor item discrimination. Questions 1, 2, 4, 7, 8, 11, 15-19, 21-23, 25, 26, and 28 are is considered weakly discriminating according to accepted standards.
3. In a study comparing response rate of correct answer and distractors, a total of 1 of 28 questions exhibits a greater response rate for the 1st distractor than for the correct response. Question 12 exhibit greater response rates for the 1st distractor than for the correct response.
4. In the same item analysis, a total of 5 of 28 questions exhibit low PBI. Questions 1, 2, 4, 8, and 13 exhibit a low PBI meaning they are considered potentially unreliable according to accepted standards.
5. Distribution of all artifact scores are centered on 26/28 and are normal exhibiting a large negative skew, or slight shift of the peak towards higher scores.
6. No comparison of dual enrollment to traditional (non-online) artifacts could be made because all sections offered during spring 2017 were offsite dual enrollment sections.
7. No comparison of online to traditional artifacts could be made because no online sections of the course were offered during spring 2017.
8. No cross-campus comparison could be made because all sections offered during spring 2017 were offsite dual enrollment sections.

9. In a longitudinal study of item analytics, questions 11 1, 13, 15, 16, 17, 20, and 21 exhibit poor performance over both terms over the course of the study (fall 2016 and spring 2017).
10. In a longitudinal study of score distribution, each term exhibits similar distribution characteristics. Both are very negatively skewed and both exhibit a small distribution at the lower end of the scoring range.

References

Ding, L. and Beichner, R. 2009. Approaches to data analysis of multiple-choice questions. Physical Review Special Topics – Physics Education Research, 5, 1-17.

Doran, R. 1980. Basic Measurement and Evaluation of Science Instruction. National Science Teachers Association, Washington, D.C., 131pp.

Suskie, L. 2004. Assessing Student Learning. Anker Publishing Co., Inc., Bolton, Massachusetts, 331 pp.