# Accounting Assessment Report
# Spring 2022

Author: Joseph F. van Gaalen, Ph.D., Asst. VP, IR, Assessment & Effectiveness

## 1 INTRODUCTION

Florida SouthWestern State College's Business Department gathers a multitude of data from various courses as assessment tools. The three courses included in this assessment report are ACG 2021 *Financial Accounting*, ACG 2071 *Managerial Accounting*, ACG 2450 *Accounting Software Applications*, and ACG 2930 *Special Topics / Capstone-Accounting*. 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 (concurrent) and non-dual enrollment students, by modality, 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, Asst. VP, IR, Assessment & Effectiveness (jfvangaalen@fsw.edu; x16965).

## 2 ACG 2021

### 2.1 ASSESSMENT ITEM ANALYSIS

Insight into the strength of the assessment questions offers information on student learning and helps to discriminate those students who 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 ACG 2021 common assessment, first administered this term, Fall 2021, consists of 44 questions in a multiple-choice and true-false format. Item difficulty for each of these questions 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 ACG 2021. The graph has a gradient such that questions calculated within the range of moderately easy to moderately difficult (see Table 1) fall in the green shaded region fading to red for questions that are deemed too easy (near 1) or too difficult (near 0). A total of 7 of 44 questions exhibit scores outside the range of what is typically defined as reasonable. Questions 12, 13, 26, 28, 31, 38, and 41 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.

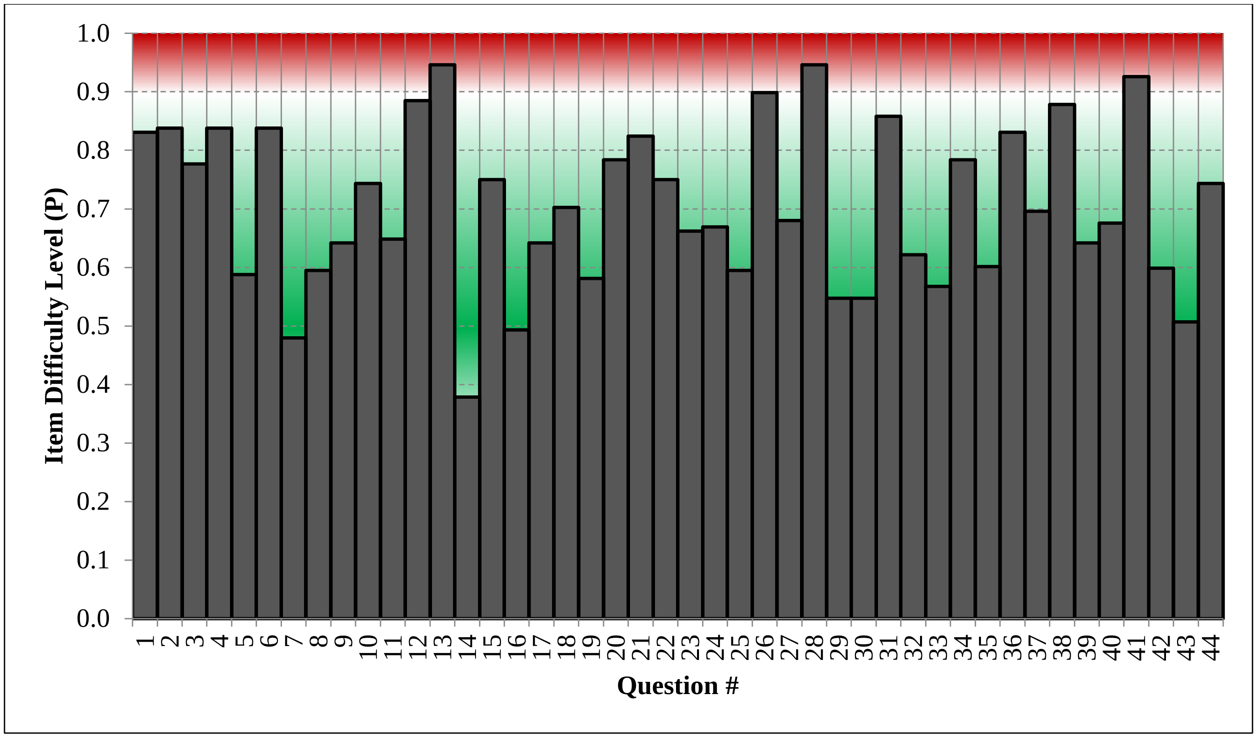


*Figure 1. Item difficulty results for ACG 2021 common assessment. Green depict moderate questions with darker green centered on the generally accepted 0.5. Lighter green to red depict questions progressively more difficult (near 0) or easier (near 1).*

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 ACG 2021. 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. Note that true/false questions are excluded from this analysis because item discrimination of this kind is not valid for a true/false question. This means questions 6, 7, 13-15, 18, 20, 23, 26, 31, 34, 37, and 40 are excluded. As a result, the only questions which exhibit poor item discrimination scores according to the literature (Doran, 1980) are questions 1, 12, 28, 38, and 41.

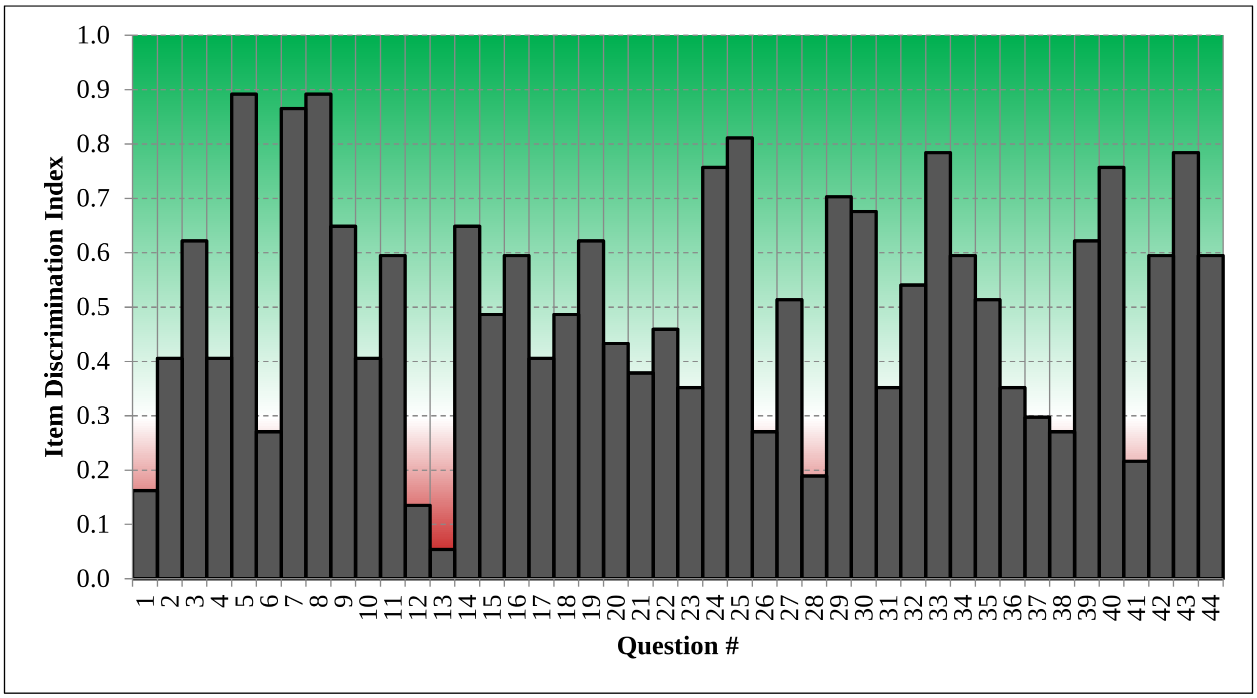


*Figure 2. Item discrimination index results for BSC 1010 common assessment. Green depict progressively stronger discriminators. White depict moderately strong (>0.3) to moderately weak (<0.3) discriminators. Red depict weak discriminators.*

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 for the common assessment. Only questions 7 and 14, true/false questions, exhibits a lower response rate for the correct response than for the 1$^{st}$ distractor.

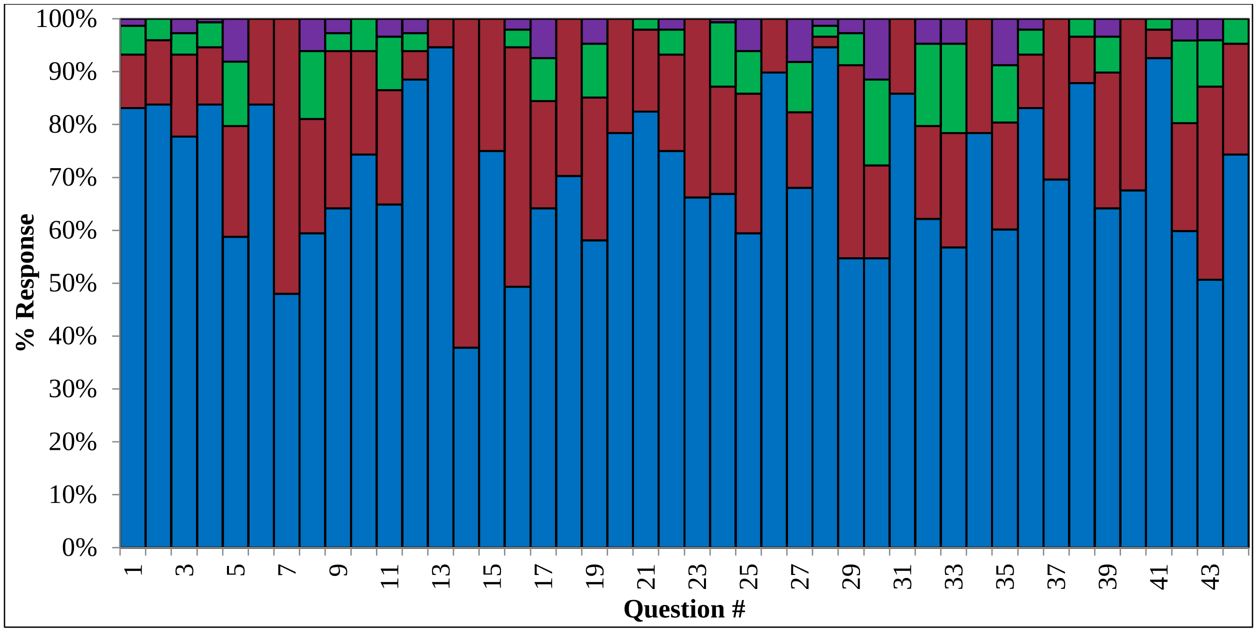


*Figure 3. Item response distribution for BSC 1010 depicting selection percentages of each response option for each question. Blue denotes correct responses, red the most common distractor, green the 2nd most, and purple the 3rd most.*

| Q | Correct | 1st Distractor | 2nd Distractor | 3rd Distractor |
|---|---|---|---|---|
| 1 | A | C | D | B |
| 2 | B | D | A | C |
| 3 | B | D | C | A |
| 4 | B | C | D | A |
| 5 | A | B | D | C |
| 6 | A | B | | |
| 7 | B | A | | |
| 8 | D | C | B | A |
| 9 | C | A | D | B |
| 10 | A | C | B | D |
| 11 | C | A | D | B |
| 12 | C | B | A | D |
| 13 | A | B | | |
| 14 | B | A | | |
| 15 | A | B | | |
| 16 | D | C | B | A |
| 17 | C | A | D | B |
| 18 | B | A | | |
| 19 | B | D | C | A |
| 20 | B | A | | |
| 21 | D | C | A | B |
| 22 | B | D | A | C |
| 23 | B | A | | |
| 24 | B | D | C | A |
| 25 | B | D | A | C |
| 26 | A | B | | |
| 27 | A | C | B | D |
| 28 | A | C | B | D |
| 29 | A | C | D | B |
| 30 | D | B | A | C |
| 31 | B | A | | |
| 32 | A | D | B | C |
| 33 | D | B | C | A |
| 34 | B | A | | |
| 35 | A | D | B | C |
| 36 | B | A | D | C |
| 37 | B | A | | |
| 38 | A | C | B | D |
| 39 | A | C | D | B |
| 40 | B | A | | |
| 41 | C | D | A | B |
| 42 | B | D | C | A |
| 43 | C | B | A | D |
| 44 | D | A | B | C |

*Table 3. Response option for corresponding correct responses and distractors for ACG 2021. Color reflect that of Figure 3 above.*

In addition to item difficulty and discrimination index, a Point Biserial Index (PBI) was calculated for each of the assessment questions. 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 4 depicts the PBI results for ACG 2021. 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. Note that true/false questions are excluded from this analysis because item discrimination of this kind is not valid for a true/false question. This means questions 6, 7, 13-15, 18, 20, 23, 26, 31, 34, 37, and 40 are excluded. As a result, only question 12 exhibits a low PBI. 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.

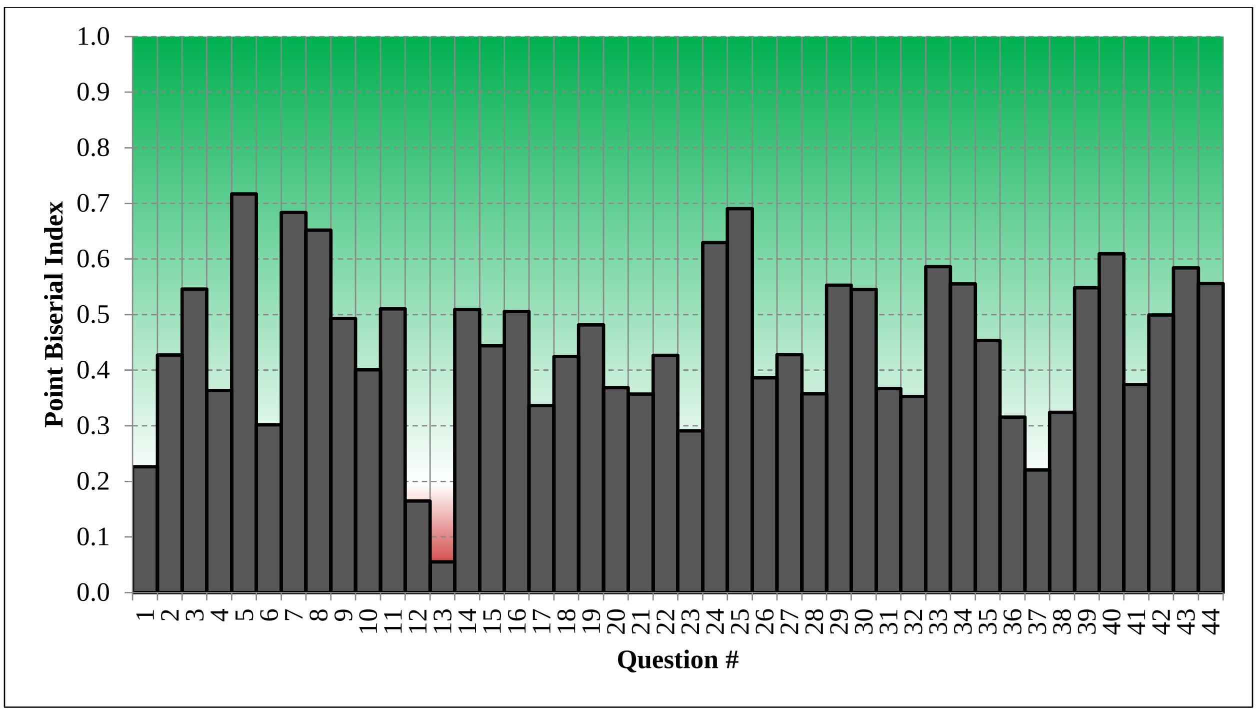


*Figure 4. Point Biserial Index (PBI) results for ACG 2021. Green shaded regions depict progressively stronger reliability. White shaded regions depict moderately strong reliability. Red shaded regions depict weak reliability.*

A full list of questions which scored poorly in each of the three item analyses is shown in Table 4. Question 12 exhibits poor scores in all three item analytics. Overall, 7/44 questions exhibit poor item difficulty scores, 5/44 questions exhibit poor item discrimination scores, and 1/44 exhibit poor PBI scores.

| Item | Item Difficulty | Item discrimination index | PBI |
|---|---|---|---|
| **Q1** | | Weak | |
| **Q12** | Too easy | Weak | Low |
| **Q13** | Too easy | | |
| **Q26** | Too easy | | |
| **Q28** | Too easy | Weak | |
| **Q31** | Too easy | | |
| **Q38** | Too easy | Weak | |
| **Q41** | Too easy | Weak | |

*Table 4. List of items that are outside the generally accepted scores of a strong multiple-choice question for ACG 2021.*

## 2.2 Learning Objectives and Descriptive Statistics

The FSW Business faculty defined one area of interest for evaluation in support of the state framework outcome. For the Spring 2022 assessment, 148 artifacts were collected for ACG 2021 from 7 of 9 course sections. There was no data available to collect in one section, and shuffled data in the other. Artifact scores are bimodal and centered on 21/44 and 41/44. The distribution exhibits a small negative skew, meaning scores are tending slightly towards higher values (Starkweather, 2010). The distribution of scores is presented in Table 5 and Figure 5.

| *Maximum score* | *44* |
|---|---|
| n | 148 |
| Max | 44 |
| Min | 13 |
| Median | 31 |
| Mode | 41 |
| Mean | 31.0 |
| Standard deviation | 8.82 |
| Skewness | -0.21 |
| Kurtosis | -1.18 |

*Table 5. Descriptive statistics for ACG 2021 common course assessment.*

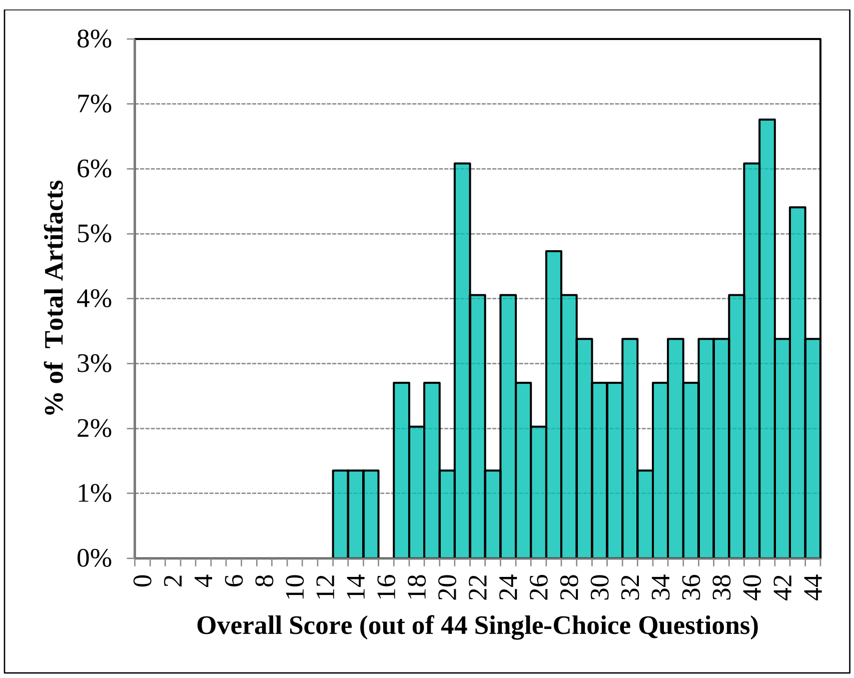


*Figure 5. Score distribution for ACG 2021 artifacts.*

## 2.3 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.

### 2.3.1 Dual Enrollment (Concurrent) to Non-Dual Enrollment Comparison

No dual enrollment (concurrent) sections of the course were run during spring 2022 so no comparison study between dual enrollment and non-dual enrollment could be completed.

### 2.3.2 Modality Comparison

During the Spring 2022 semester, 100 total asynchronous online artifacts were collected, along with 20 live online, 0 flex, 0 blended, and 28 traditional. A comparison of basic statistics is provided in Table 6. Traditional artifacts mean scores are the highest at 39.9, followed by asynchronous online at 29.2, and live online at 27.7. A one-way analysis of variance was used to compare means by modality. Results of the ANOVA exhibit a statistically significant difference between sites (see Table 7). Therefore, we can reject the null hypothesis that the mean scores at each site are equal to each other and we can conclude with a 95% confidence that the differences in scores are not solely due to chance. A distribution of scores by modality is shown in Figure 6.

| | *Traditional* | *Async Online* | *Live Online* | *Flex* | *Blended* |
|---|---|---|---|---|---|
| n | 28 | 100 | 20 | ~ | ~ |
| Max | 44 | 44 | 42 | ~ | ~ |
| Min | 26 | 13 | 15 | ~ | ~ |
| Median | 41 | 29 | 25.5 | ~ | ~ |
| Mode | 41 | 21 | 22 | ~ | ~ |
| Mean | 39.9 | 29.2 | 27.7 | ~ | ~ |
| Standard deviation | 3.84 | 8.38 | 8.41 | ~ | ~ |
| Skewness | -2.04 | 0.03 | 0.27 | ~ | ~ |
| Kurtosis | 5.49 | -0.98 | -1.22 | ~ | ~ |

*Table 6. Comparison of statistics by modality.*

| **Source of Variation** | **Sum of squared differences** | **df** | **Mean Squares** | **$F_{obs}$** | **p-value** | **$F_{crit}$** |
|---|---|---|---|---|---|---|
| Between Sites | 2737.8 | 2 | 1368.9 | 22.83 | $2.40 \times 10^{-9}$ | 3.06 |
| Within Sites | 8695.1 | 145 | 60.0 | | | |
| Total | 11,432.9 | 147 | | | | |

*Table 7. Results of one-way ANOVA of mean scores in each modality for ACG 2021.*

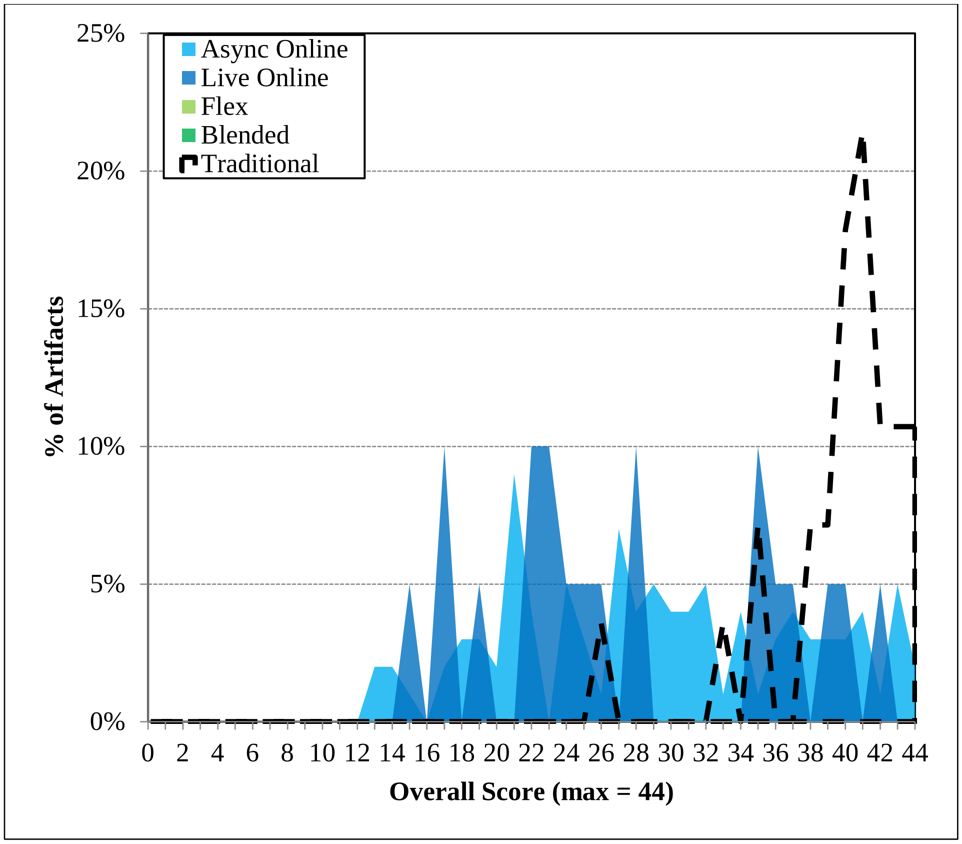


*Figure 6. Score distribution by modality.*

#### 2.3.3 Comparison by Campus/Site

Only one traditional site reported data so no comparison across sites could be completed.

### 2.4 Longitudinal Study

A longitudinal study for this project will begin following the Fall 2022 term.

# 3 ACG 2071

### 3.1 Learning Objectives and Descriptive Statistics

Using a common course assessment, the FSW Business faculty identified the Section Test 2 (Chapters 5, 6, & 7) from Managerial Accounting course (ACG 2071) will be used for this assessment method. The benchmark of 70% of students will illustrate a proficiency of 70% or higher within this assessment during the 2021-2022 academic year. Discussion during the April 2021 and August 2021 School of Business and Technology Business Department breakout meeting finalized the decision regarding the assessment tool to be used for the learning outcome.

For the Spring 2022 assessment, 107 artifacts were collected for ACG 2071 from 3 of 5 course sections. Artifact scores are centered on 85/100 (Table 8). The goal of 70% of artifacts scoring 70% or higher was met, at 75.7%.

| Maximum score | 100 |
|---|---|
| n | 107 |
| Max | 100 |
| Min | 25 |
| Mode | 85 |
| Mean | 77.5 |
| % Achieving 70% | 75.7% |

*Table 8. Descriptive statistics for ACG 2071 common course assessment.*

## 3.2 Exploratory Analysis and Significance Testing

Multiple comparisons of artifact scores across varying formats, campuses, and student types were made, where possible, 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.

### 3.2.1 Dual Enrollment (Concurrent) to Non-Dual Enrollment Comparison

No dual enrollment (concurrent) sections of the course were run during spring 2022 so no comparison study between dual enrollment and non-dual enrollment could be completed.

### 3.2.2 Modality Comparison

During the Spring 2022 semester, 51 total asynchronous online artifacts were collected, along with 0 live online, 26 flex, 0 blended, and 30 traditional. A comparison of basic statistics is provided in Table 2. Flex exhibits the highest achievement at 88.5%, followed by asynchronous online, at 78.4%, and then traditional, at 60.0%. There is a statistically significant difference between flex and traditional, but no difference involving asynchronous online (i.e., neither statistically significantly different to traditional nor flex).

| | *N* | *Mean* | *% Achieving 70% or Higher* |
|---|---|---|---|
| *Asynchronous Online* | 51 | 78.8 | 78.4% |
| *Live Online* | 0 | ~ | ~ |
| *Flex* | 26 | 84.6 | 88.5% |
| *Blended* | 0 | ~ | ~ |
| *Traditional* | 30 | 69.0 | 60.0% |

*Table 9. Percent achieving 70% or higher score by modality.*

### 3.2.3 Comparison by Campus/Site

Only one traditional site was run, so no comparison by site could be completed.

## 3.3 Longitudinal Study

A longitudinal study for this project will begin following the Fall 2022 term.

# 4 ACG 2450

For the Spring 2022 assessment no data was able to be collected. There was no data available to collect. The expected “Comprehensive Exam" or old assessment tool "Quickbooks Application Ch 8 Rubric" was available in the Learning Management System (LMS), Canvas. A final exam was administered through an external online resource.

# 5 Conclusions

FSW's Business Department has employed a common assignment for courses as assessment tools. The three courses included in this assessment report are ACG 2021 *Financial Accounting*, ACG 2071 *Managerial Accounting*, and ACG 2450 *Accounting Software Applications*. The results are intended to provide a baseline achievement moving forward.

## 5.1 ACG 2021

A drill-down of ACG 2021 results are as follows:

1. In an item analytics study, a total of 7 of 44 questions exhibit scores outside the range of what is typically defined as reasonable. Questions 12, 13, 26, 28, 31, 38, and 41 exhibit item difficulty categorized as 'too easy' according to accepted standards.
2. In an item analytics study, in questions where this study is valid (multiple choice questions, not true/false questions), questions 1, 12, 28, 38, and 41 are categorized as weak discriminators according to accepted standards.
3. In a study of distractors, only questions 7 and 14, a true/false question, exhibits a lower response rate for the correct response than for the 1st distractor.
4. In an item analytics study, in questions where this study is valid (multiple choice questions, not true/false questions), only question 12 exhibits a low PBI according to accepted standards.
5. For the Spring 2022 assessment, 148 artifacts were collected for ACG 2021 from 7 of 9 course sections. There was no data available to collect in one section, and shuffled data in the other. Artifact scores are bimodal and centered on 21/44 and 41/44. The distribution exhibits a small negative skew, meaning scores are tending slightly towards higher values.
6. No dual enrollment (concurrent) sections of the course were run during spring 2022 so no comparison study between dual enrollment and non-dual enrollment could be completed.
7. During the Spring 2022 semester, 100 total asynchronous online artifacts were collected, along with 20 live online, 0 flex, 0 blended, and 28 traditional. Traditional artifacts mean scores are the highest at 39.9, followed by asynchronous online at 29.2, and live online at 27.7. Results of the ANOVA exhibit a statistically significant difference between sites.
8. Only one traditional site reported data so no comparison across sites could be completed.
9. A longitudinal study for this project will begin following the Fall 2022 term.

## 5.2 ACG 2071

A drill-down of ACG 2071 results are as follows:

1. For the Spring 2022 assessment, 107 artifacts were collected for ACG 2071 from 3 of 5 course sections. Artifact scores are centered on 85/100. The goal of 70% of artifacts scoring 70% or higher was met, at 75.7%.
2. No dual enrollment (concurrent) sections of the course were run during spring 2022 so no comparison study between dual enrollment and non-dual enrollment could be completed.
3. During the Spring 2022 semester, 51 total asynchronous online artifacts were collected, along with 0 live online, 26 flex, 0 blended, and 30 traditional. Flex exhibits the highest achievement at 88.5%, followed by asynchronous online, at 78.4%, and then traditional, at 60.0%. There is a statistically significant difference between flex and traditional, but no difference involving asynchronous online (i.e., neither statistically significantly different to traditional nor flex).
4. Only one traditional site was run, so no comparison by site could be completed.

## 5.3 ACG 2450

For the Spring 2022 assessment no data was able to be collected. There was no data available to collect. The expected “Comprehensive Exam" or old assessment tool "Quickbooks Application Ch 8 Rubric" was available in the Learning Management System (LMS), Canvas. A final exam was administered through an external online resource.

# 6 REFERENCES


Cohen, J. 1988. Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Earlbaum Associates, Hillsdale, NJ.

Davis, J.C. 1973. Statistics and Data Analysis in Geology. John Wiley & Sons, New York, New York, 564 pp.

Lipsey, M.W. and Wilson, D.B. 1993. The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. American Psychologist, 48, 1181-1209.

McDonald, J.H. 2009. Handbook of Biological Statistics (2nd ed.). Sparky House Publishing, Baltimore, Maryland.

Rosenthal, R. and Rosnow, R.L. 1991. Essentials of behavioral research: Methods and data analysis (2nd ed.). McGraw Hill, New York, NY.

Starkweather, J. D. 2010. Introduction to Statistics for the Social Sciences. In: Research and Statistical Support. Retrieved from http://www.unt.edu/rss/class/Jon/ISSS_SC/.

Wilkinson, L. 1999. APA Task Force on Statistical Inference. Statistical Methods in Psychology Journals: Guidelines and Explanations. American Psychologist 54 (8), 594–604.