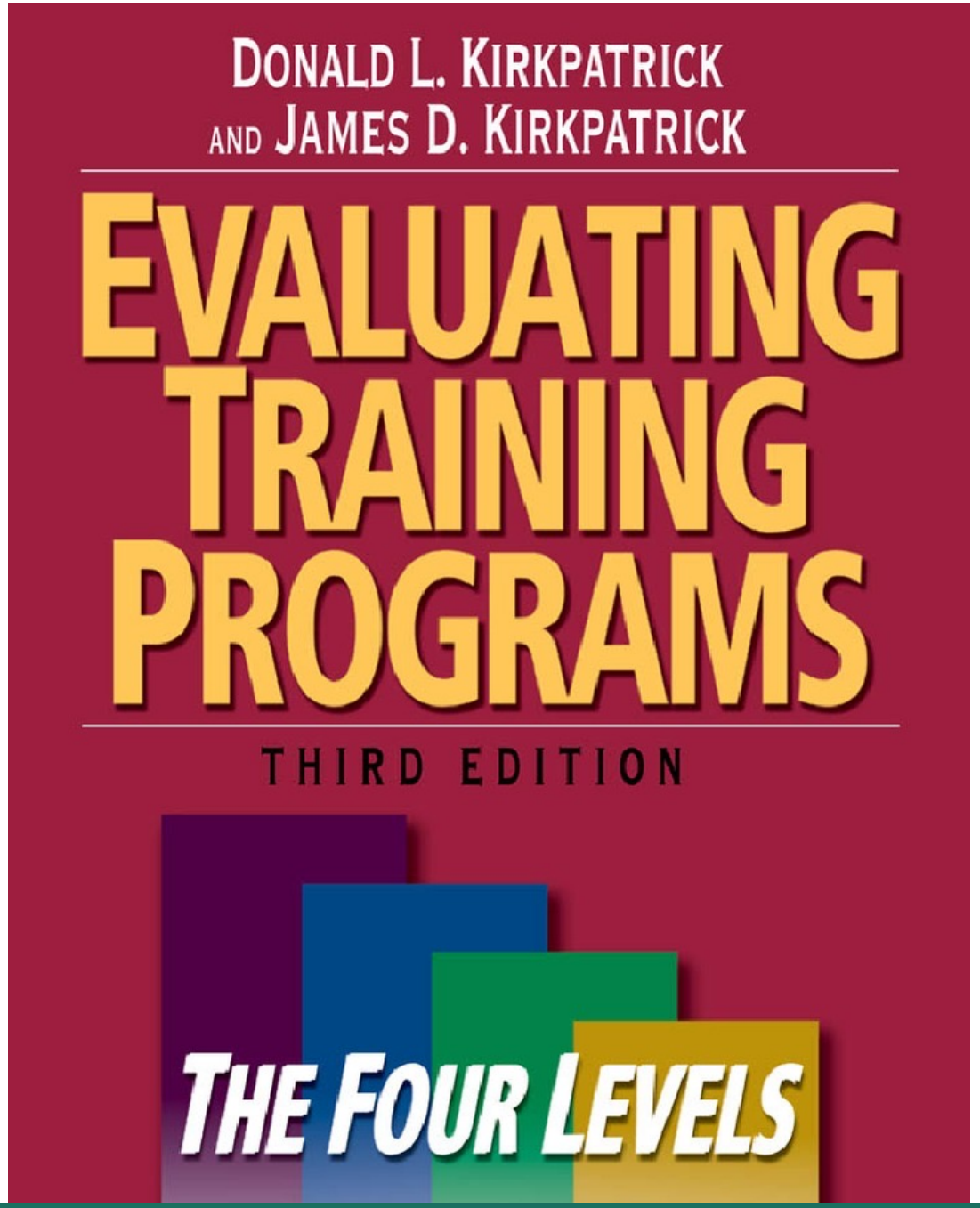
# EVALUATING A PHARMACEUTICAL CALCULATIONS COURSE USING THE KIRKPATRICK MODEL



# Learners will be able to:

- Discuss how Kirkpatrick's Four Levels of Evaluation provides a framework for evaluating a course.
- List at least four sources of data that can be used to evaluate a course.
- Critique a course evaluation using key elements of Kirkpatrick's Four Levels of Evaluation as criteria.
- Utilize a stepwise process to collect course data to defend course adjustments.

# Presenters

**Kimberly K. Daugherty, Pharm.D, Ph.D.**
Associate Dean of Academic Affairs and Assessment
Sullivan University College of Pharmacy and Health Sciences (SUCOPHS)

**Sarah Raake, PharmD, MSEd, BCACP, LDE**
Associate Professor / Director of Instructional Effectiveness
SUCOPHS

**Ben Stephens, M.Ed.**
Instructional Support Specialist
SUCOPHS

# About us: Sullivan University College of Pharmacy & Health Sciences

- Private institution in Louisville, KY
- Contains 2-year Master's level Physician Assistant (PA) program and 3-year Doctor of Pharmacy (PharmD) program
- Year-round, quarter-based system
- No large assessment office
- Faculty and course coordination self-sufficiency is a must

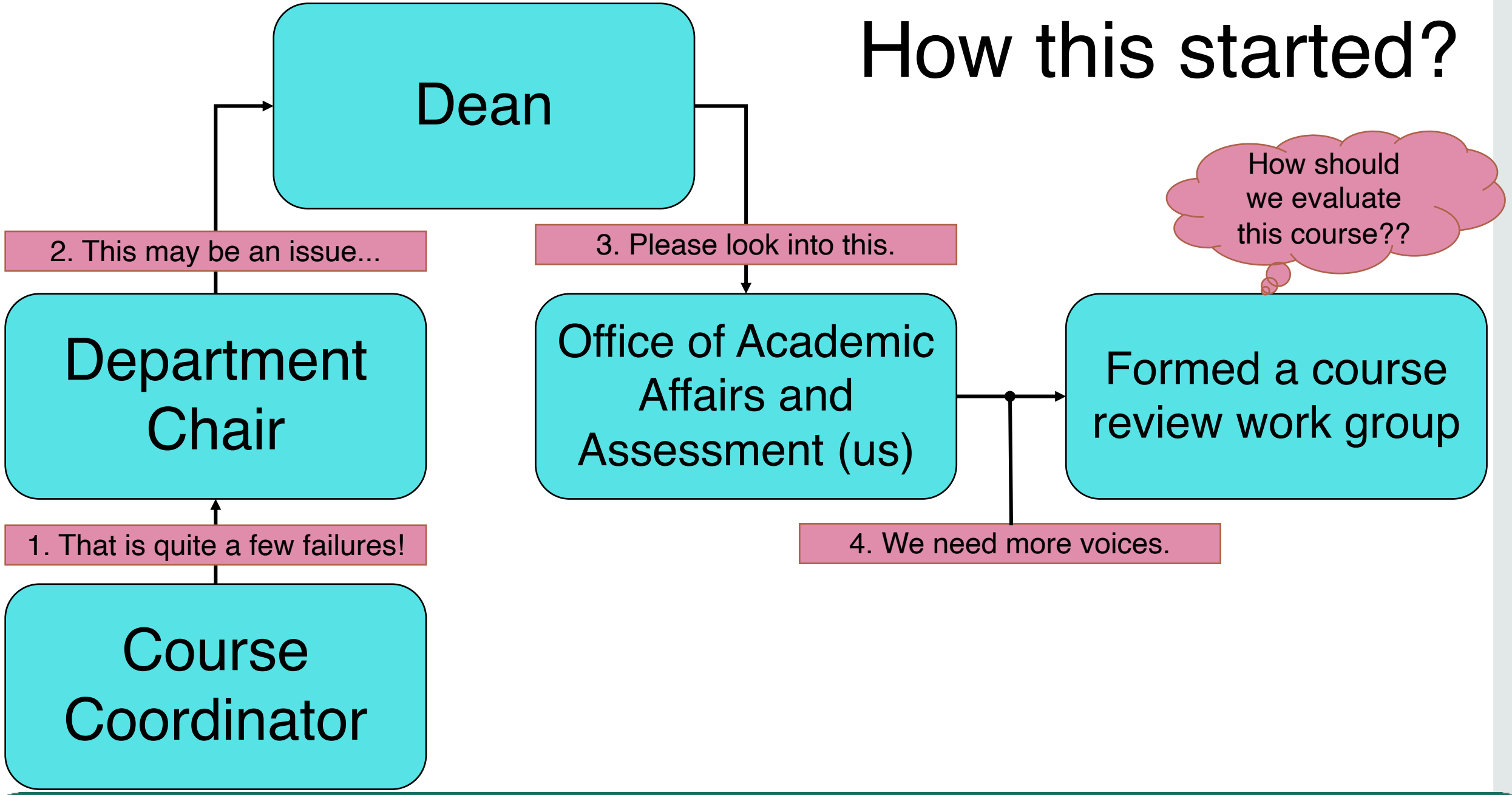
# The Situation

Several students in PharmD program were failing Pharmaceutical Calculations and were struggling on the NAPLEX (licensure exam) in this area.

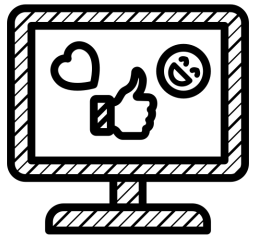
Pharmaceutical Calculations is essentially applying algebraic principles to story problems.

Many of our students had taken higher level math courses but were still struggling.

# How this started?



# Overview of Kirkpatrick

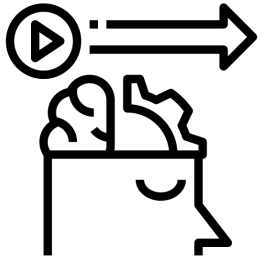


Created by Valter Bispo from the Noun Project

Level 1: Reaction

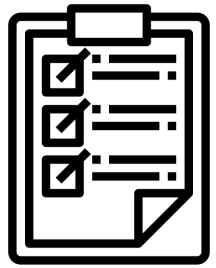
Created by Adrien Coquet from the Noun Project

Level 2: Learning



Created by Nithinan Tatal

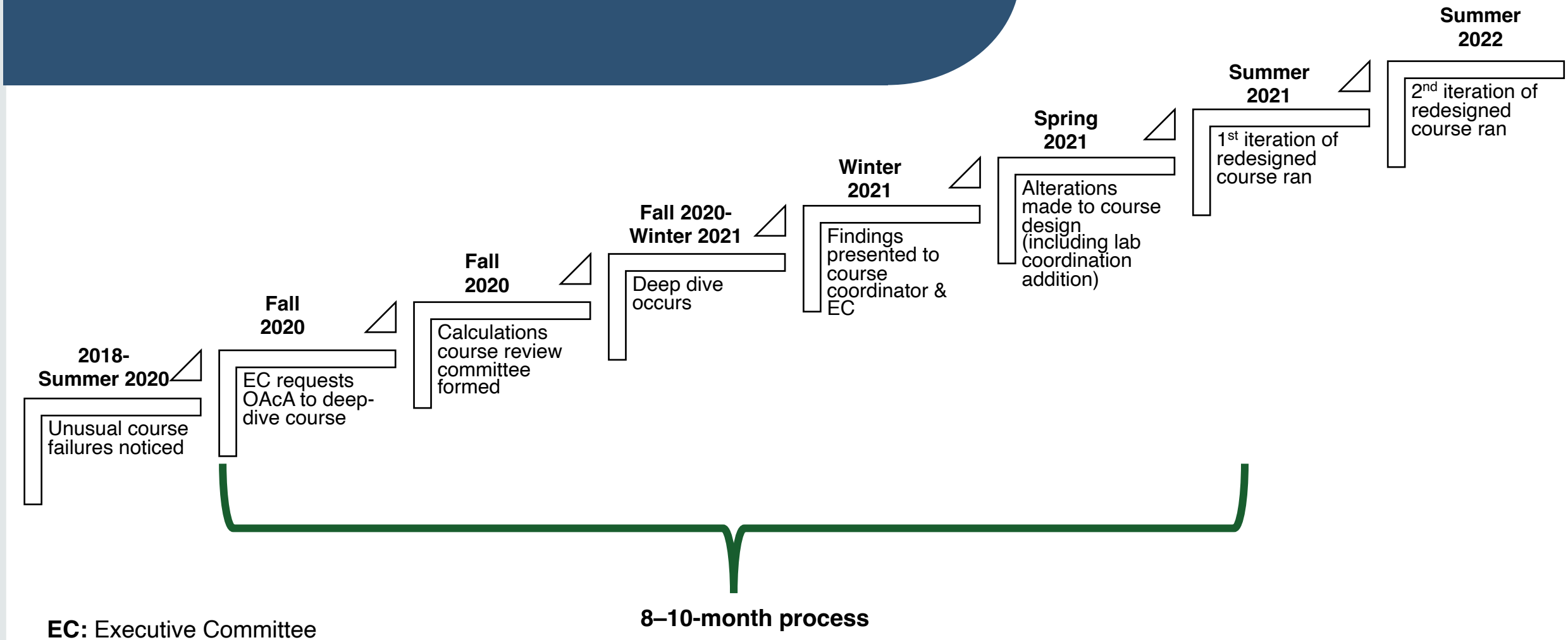
Level 3: Behavior



Created by Template

Level 4: Results

# Timeline




**EC:** Executive Committee
**OAcA:** Office of Academic Affairs & Assessment

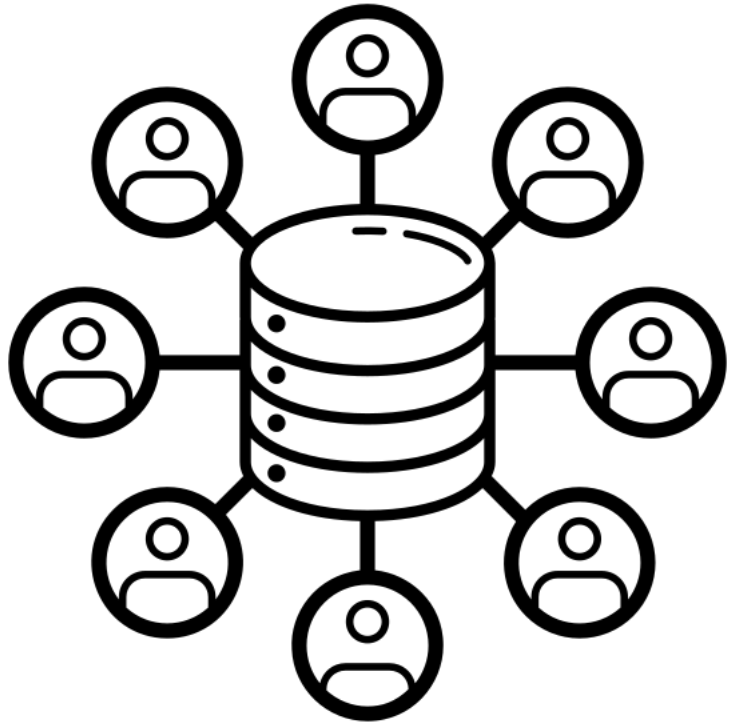
# Level One – Reaction

| Kirkpatrick Guideline | Our Process |
|---|---|
| Determine what you want to find out | Survey approved by administration |
| Design a form that will quantify reactions | LMS (Blackboard) / Qualtrics supports this |
| Encourage written comments and suggestions. | Open response questions were included |
| Get 100 percent immediate response | Completion of the surveys were tied to next quarters' grades |
| Get honest responses. | Responses were anonymous in LMS; Qualtrics confidential |
| Develop acceptable standards | Typically compared to other course scores |
| Measure reactions against standards and take appropriate action | Committee was formed to address concerns |
| Communicate reactions as appropriate. | Course coordinator and chair were kept apprised |

# Level One – Data Sources

We examined:

- Survey data from the previous 3 years



Created by Vectors Point
from the Noun Project

# The Survey

Likert Questions – Strongly Disagree, Disagree, Agree, Strongly Agree

1. The course addressed the topics outlined in the syllabus.
2. The instructor(s) followed the syllabus (including revisions communicated to the class).
3. The course was organized and flowed well.
4. In addition to slides, the other learning materials, if used (textbooks, additional reference books, videos, weblinks, etc.) were helpful to my understanding of the course.
5. The assessments (exams, quizzes, assignments, etc.) were representative of course content.
6. I was provided with adequate and timely feedback on my performance and course issues.
7. Overall, I would rate this course as a high-quality course.

Open Response

8. List any strengths of this course:
9. List ways the course be improved to help your learning experience.

# Survey Results

## The course was organized and flowed well.

| | 2018 (N=63) | 2019 (N=47) | 2020 (N=45) |
|---|---|---|---|
| Strongly Agree/Agree | 85.7% | 98% | 73.3% |
| Disagree | 6.35% | 0% | 20% |
| Strongly Disagree | 7.94% | 2.13% | 6.67% |

# Survey Results

The assessments (exams, quizzes, assignments, etc.) were representative of course content.

| | 2018 (N=63) | 2019 (N=47) | 2020 (N=45) |
|---|---|---|---|
| Strongly Agree/Agree | 69.8% | 93.7% | 82.2% |
| Disagree | 20.6% | 4.3% | 11.1% |
| Strongly Disagree | 9.5% | 2.1% | 4.4% |

# Survey Results

## Overall, I would rate this course as a high-quality course.

| | 2018 (N=63) | 2019 (N=47) | 2020 (N=45) |
|---|---|---|---|
| Strongly Agree/Agree | 73% | 97.9% | 66.7% |
| Disagree | 19.1% | 2.1% | 20% |
| Strongly Disagree | 7.9% | 0% | 11.1% |

# Survey Results
## (Open Response)

List ways the course could be improved to help your learning experience. (A few themes)

| 2018 | 2019 | 2020 |
| --- | --- | --- |
| Extra practice problems | Need more practice problems | Need more practice problems |
| Incorrect arithmetic / errors | Incorrect arithmetic / errors | Too fast paced |
| Exams need to connect to classwork | Homework was easier than assessments | Poor alignment between HW, quizzes, and exams |
| Students need to be taught systematic approach for solving problems | Professor to student communication | Professor to student communication |

# Level One – Findings

- Students found practice problems beneficial and wanted additional practice problems.
- Students were concerned about arithmetic errors with lectures and homework questions.
- Students perceived the instructional practice problems, the quizzes, and the summative assessments were not written at the same level of complexity.
- Some students felt that there was not a clear organization or structure to the course.

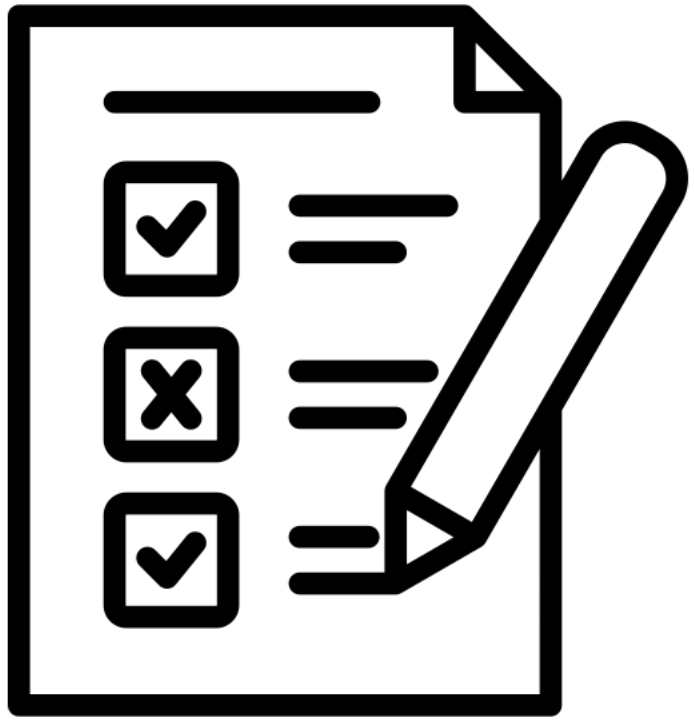
# Level Two – Learning

| Kirkpatrick Guideline | Our Process |
|---|---|
| Use a control group in practical | |
| Evaluate knowledge, skills, and/or attitudes both before and after the program. | |
| Use a paper-and-pencil test to measure knowledge and attitudes. | LMS quizzes were used as formative assessments throughout the course. |
| Use a performance test to measure skills | Examsoft assessment were provided throughout the course. |
| Get a 100 percent response. | All students completed the assessments. |
| Use the results of the evaluation to take appropriate action. | Exam data was reviewed to make changes in exams and course. |

# Level Two – Data Sources

We examined:

- Summative assessment data from our assessment platform, Examsoft



Created by Eko Purnomo
From the Noun Project

# Exam 1 Overview – 2018-2020

## 2018

### ASSESSMENT PERFORMANCE

| 85% | 29% | 100% |
|---|---|---|
| Average Score (38.2/45) | Low Score (13/45) | High Score (45/45) |

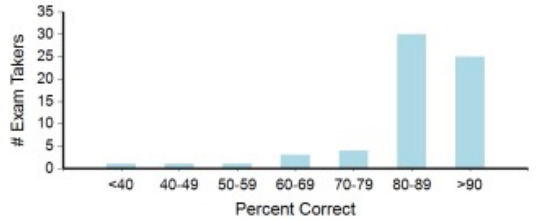
**Assessment Score Reliability (KR-20)**

0.0 — 0.83 — 1.0

POOR | SATISFACTORY | GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2019

### ASSESSMENT PERFORMANCE

| 82% | 24% | 106% |
|---|---|---|
| Average Score (31.2/38) | Low Score (9.2/38) | High Score (40.2/38) |

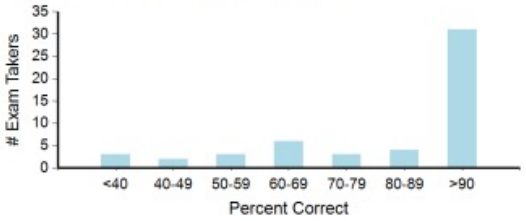
**Assessment Score Reliability (KR-20)**

0.0 — 0.91

POOR | SATISFACTORY | GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2020

### ASSESSMENT PERFORMANCE

| 77% | 7% | 100% |
|---|---|---|
| Average Score (31.5/41) | Low Score (3/41) | High Score (41/41) |

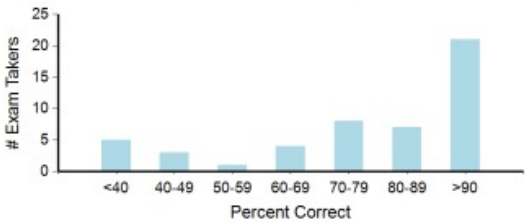
**Assessment Score Reliability (KR-20)**

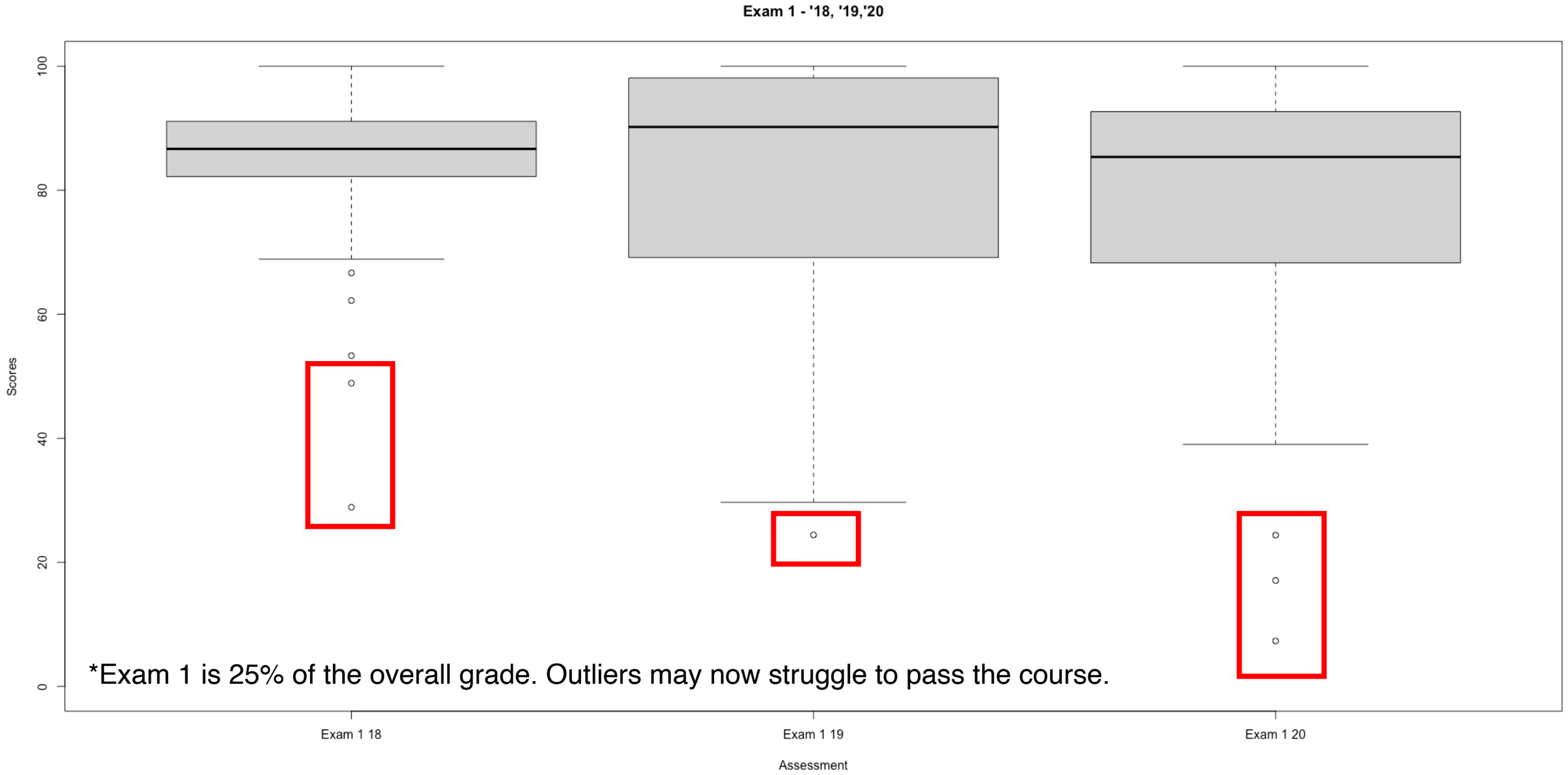
0.0 — 0.93

POOR | SATISFACTORY | GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**






*Exam 1 is 25% of the overall grade. Outliers may now struggle to pass the course.

# Exam 2 Overview – 2018-2020

## 2018

### ASSESSMENT PERFORMANCE

| 77% | 25% | 100% |
|---|---|---|
| Average Score (30.9/40) | Low Score (10/40) | High Score (40/40) |

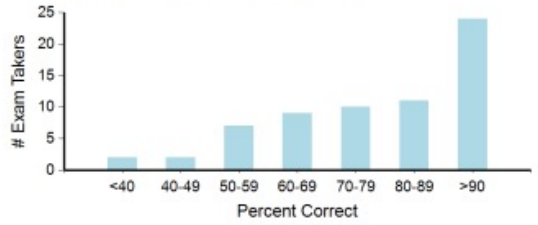
**Assessment Score Reliability (KR-20)**

0.0 0.89 1.0

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2019

### ASSESSMENT PERFORMANCE

| 83% | 47% | 107% |
|---|---|---|
| Average Score (29.1/35) | Low Score (16.4/35) | High Score (37.4/35) |

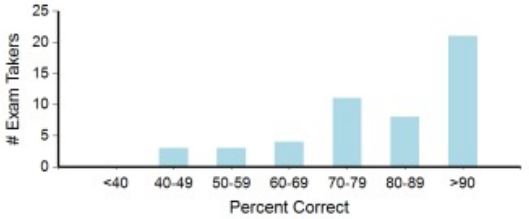
**Assessment Score Reliability (KR-20)**

0.0 0.85 1.0

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2020

### ASSESSMENT PERFORMANCE

| 79%* | 29%* | 103%* |
|---|---|---|
| Average Score (26.8/34)* | Low Score (10/34)* | High Score (35/34)* |

*Bonus points were available on this assessment and were calculated into the score.

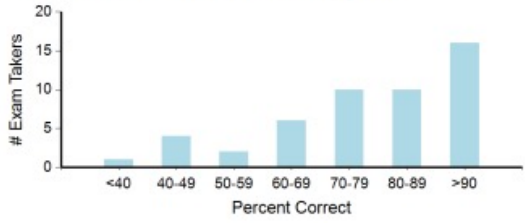
**Assessment Score Reliability (KR-20)**

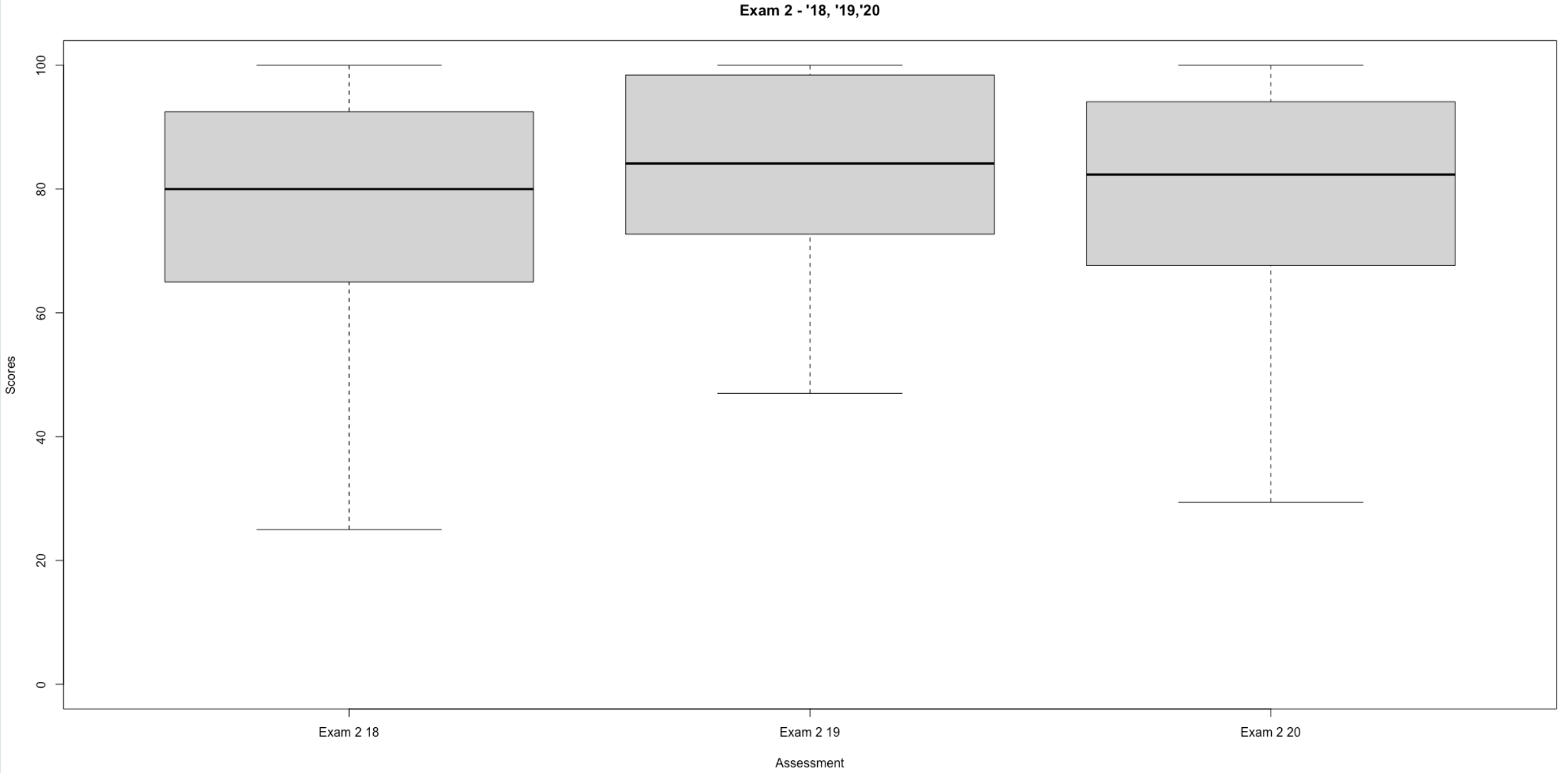
0.0 0.87 1.0

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




Exam 2 - '18, '19,'20
Scores
0
20
40
60
80
100
Exam 2 18
Exam 2 19
Exam 2 20
Assessment

# Final Exam Overview – 2018-2020

## 2018

**ASSESSMENT PERFORMANCE**

80% Average Score (40.1/50)

32% Low Score (16/50)

98% High Score (49/50)

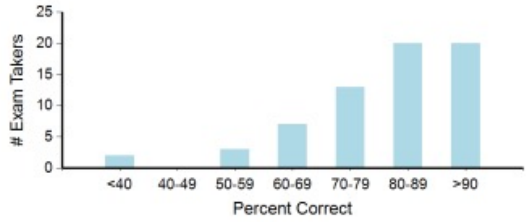
**Assessment Score Reliability (KR-20)**

0.0 0.87 1.0

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2019

**ASSESSMENT PERFORMANCE**

83% Average Score (33.2/40)

55% Low Score (22/40)

100% High Score (40/40)

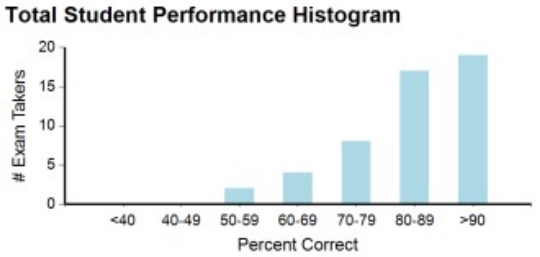
**Assessment Score Reliability (KR-20)**

0.0 0.78 1.0

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2020

**ASSESSMENT PERFORMANCE**

81%* Average Score (32.4/40)*

18%* Low Score (7/40)*

103%* High Score (41/40)*

*Bonus points were available on this assessment and were calculated into the score.

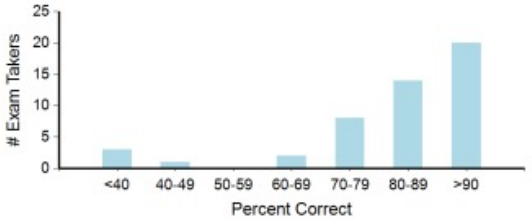
**Assessment Score Reliability (KR-20)**

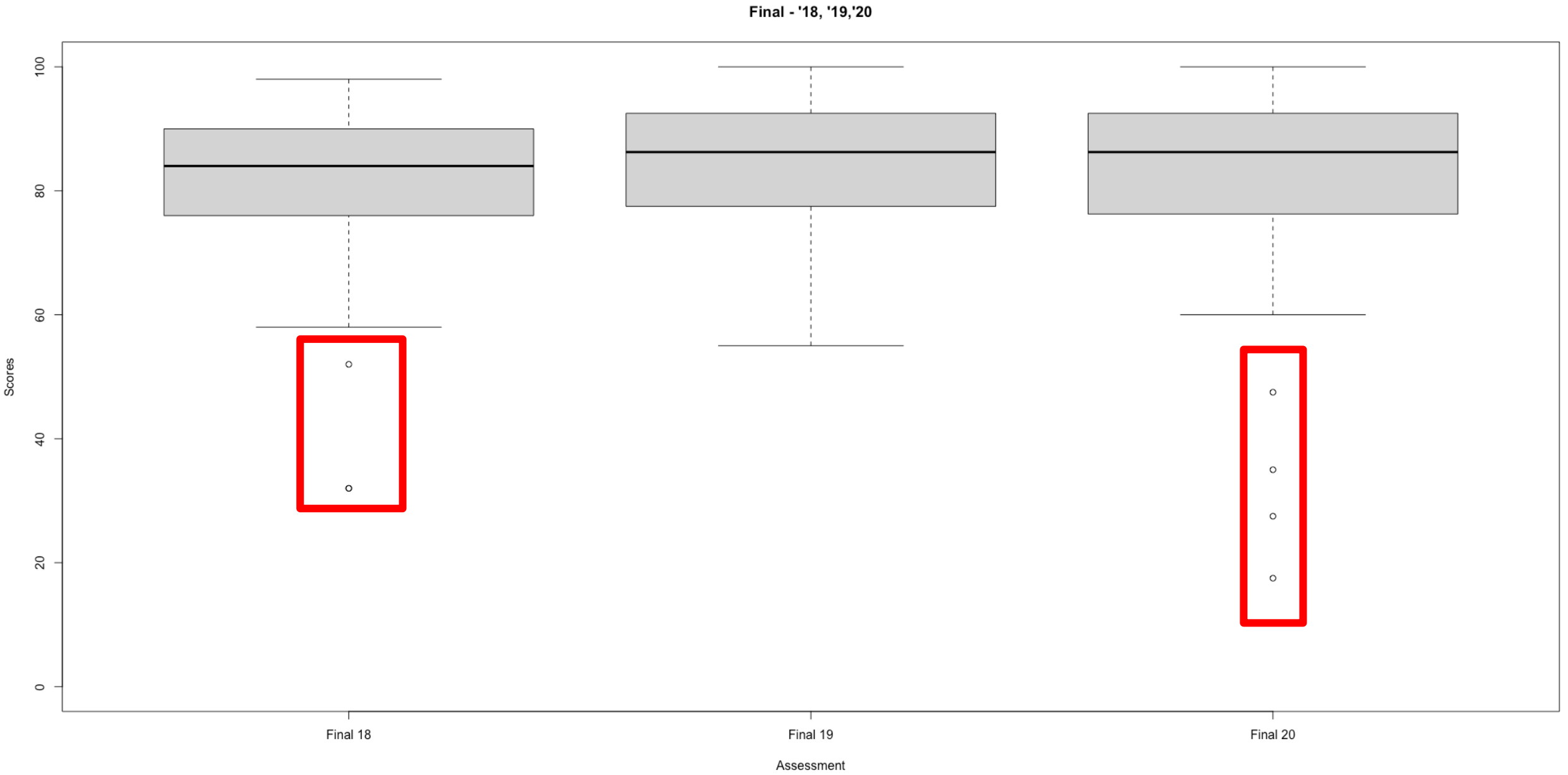
0.0 0.9

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




Final - '18, '19,'20
Scores
0
20
40
60
80
100
Final 18
Final 19
Final 20
Assessment

# Level Two – Findings

- Most students were performing well. The mean score on assessments was typically around 80% or above.
- There was a wide range of scores.
- There were a few students who scored low enough on the exams to be considered outliers. A student who scores 30% on exam 1 would need to score an average of 83% on the remaining assignments / assessments to pass the course.
- "Either students were getting it, or they didn’t."

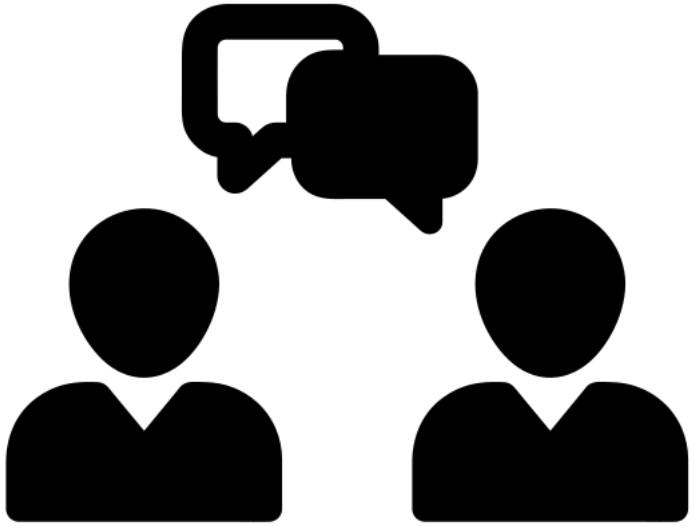
# Level Three – Behavior

| Kirkpatrick Guideline | Our Process |
|---|---|
| Use a control group if practical. | |
| Allow time for behavior change to take place. | We waited 2 months to conduct the focus group. |
| Evaluate both before and after the program if practical. | |
| Survey and/or interview one or more of the following: trainees, their immediate supervisor, their subordinates, and other who often observe their behavior | A focus group was conducted a few weeks after the course completed. |
| Get 100 percent response or a sampling | We choose a sampling of high and low performers. |
| Repeat the evaluation at appropriate times. | |
| Consider cost versus benefits | |

# Level Three – Data Sources

We examined:

- Feedback from a focus groups
- Admission and course performance data to look for early warning signs of possible course failure

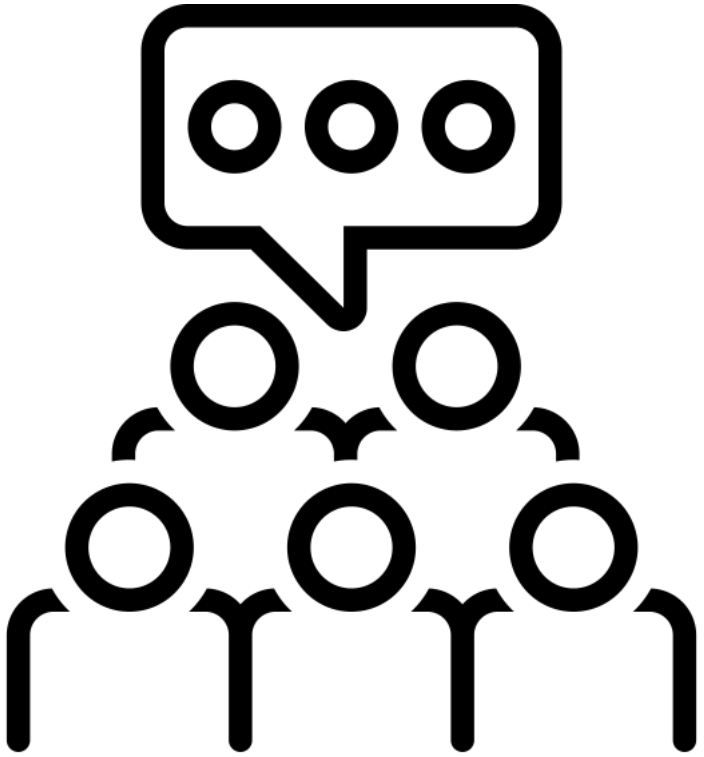


Created by shashank singh
from noun project

# Focus Group

The focus group:

- was led by three faculty members not associated with the course.
- was comprised of students who performed well in the course and students who struggled.



Created by WEBTECHOPS LLP
From Noun Project

# Focus Group Findings

## Homework / Practice

- Arithmetic mistakes within the exiting practice problems
- Unable to practice specific questions or sort questions by topic
- Students found the practice problems helpful

## Summative Assessment

- Difficulty was inconsistent between homework, quizzes, and tests
- Not enough time on exams
- Students wanted more time between lectures and assessment

## Course

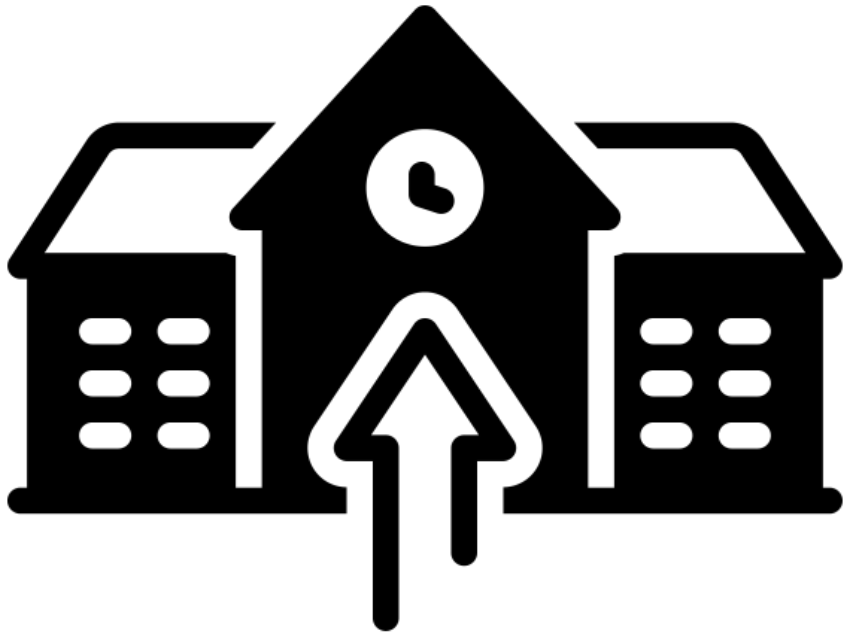
- Too fast paced
- There were issues with team members in some of the groups
- The overall grade incorporated multiple assessments

# Course Data Reviewed

- Admissions data
  - Correlations with Pre-admit Math GPA
  - Overall Pre-req GPA
  - Total math hours
- Course data
  - Exam scores to individual course assignment grades
  - Exam scores to Blackboard review activities
  - Review of exam questions from Exam 1

# Admission Correlation data

- Pre-admit math GPA = 0.315 (weak correlation)
- Overall pre-rea GPA = 0.27 (weak correlation)
- Total math hours = -0.12 (weak negative correlation)



Created by WEBTECHOPS LLP
From the Noun Project

# Course Data

- Individual assignments
  - Exam 1 score and Week 2 individual assessment: 0.73 (strong correlation)
  - Exam 1 score and average of Week 2-4 individual assessment: 0.61 (moderate correlation)

- Blackboard review activities
  - Exam 1 score vs Week 1-2 review (n=25): 0.72 (strong correlation)
  - Exam 1 score vs Week 1-3 review (n=20): 0.75 (strong correlation)

- ExamSoft question review for Exam 1
  - Exam discriminated between the top and bottom 27%
  - In review of questions missed by bottom 27%

    Lots of wordy word problems

    Several multi-step problems

# Monitoring through Intentional Interleaving

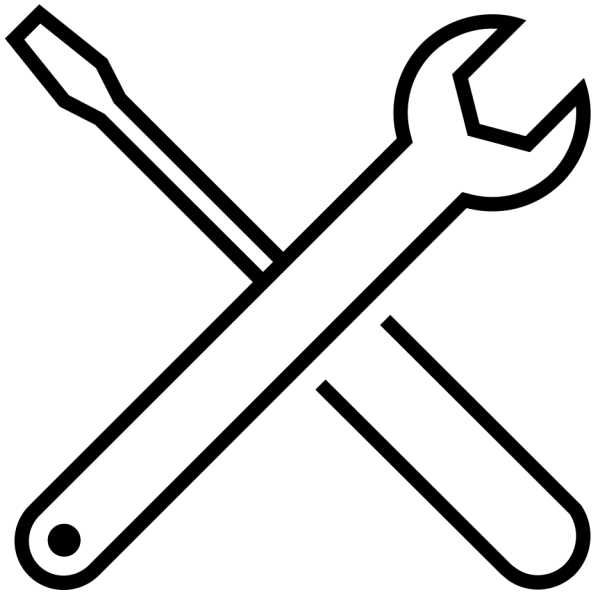
- Calculations needed interleaved into the curriculum
- Subsection of calculations items included in every final exam of lab sequence
- Study materials/reviews posted for retrieval practice
- Continued assessment of retention

# Course Adjustments

## Big Picture

This course was working for most students.

How could we support struggling students without making major changes? The course needed fine tuning.



# Detailed Course Adjustments

| Finding | Adjustment |
| --- | --- |
| Students felt additional practice problems were helpful and wanted more. | Created a script in R to generate multiple versions of existing calculations problems. |
| Students wanted additional communication with the course coordinator. | A lab coordinator was added to the course. |
| A small percentage of students clearly did not learn the material. | A lab coordinator was added to expand office hours and additional practice problems were provided. |
| Students wanted more time between instruction and assessment. | The course scheduled was adjusted. |
| Early assessments within the course were correlated with overall performance in the course. | Coordinators made sure to reach out to students who failed early assessments. |
| Some of the assessment items were excessively wordy. | Coordinators reviewed test items. |

# Level Four – Results

| Kirkpatrick Guideline | Our Process |
|---|---|
| Use a control group if practical. | |
| Allow time for results to be achieved. | Changes have been run for 1 year; continue to monitor. |
| Measure both before and after the program if practical. | Data was collected prior to change and after changes were implemented. |
| Repeat the measurement at appropriate times. | Exams and course evaluations were administered along the usual schedule. |
| Consider cost versus benefits | Course data was followed to ensure addition of the lab coordinator was helpful. Practicing pharmacists would NOT be assessed. |
| Be satisfied with evidence if proof is not possible. | KEY POINT! We needed to show “upper management the training was worthwhile” at the end of the course. |

# Survey Results - Update

## The course was organized and flowed well.

| | 2018 (N=63) | 2019 (N=47) | 2020 (N=45) | 2021 (N=40) |
|---|---|---|---|---|
| Strongly Agree/Agree | 85.7% | 98% | 73.3% | 82.5% |
| Disagree | 6.35% | 0% | 20% | 17.5% |
| Strongly Disagree | 7.94% | 2.13% | 6.67% | 0% |

# Survey Results - Update

The assessments (exams, quizzes, assignments, etc.) were representative of course content.

| | 2018 (N=63) | 2019 (N=47) | 2020 (N=45) | 2021 (N=40) |
|---|---|---|---|---|
| Strongly Agree/Agree | 69.8% | 93.7% | 82.2% | 100% |
| Disagree | 20.6% | 4.3% | 11.1% | 0% |
| Strongly Disagree | 9.5% | 2.1% | 4.4% | 0% |

# Survey Results

## Overall, I would rate this course as a high-quality course.

| | 2018 (N=63) | 2019 (N=47) | 2020 (N=45) | 2021 (N=40) |
|---|---|---|---|---|
| Strongly Agree/Agree | 73% | 97.9% | 66.7% | 95% |
| Disagree | 19.1% | 2.1% | 20% | 2.5% |
| Strongly Disagree | 7.9% | 0% | 11.1% | 2.5% |

# Exam 1 Overview – 2018-2021

## 2018

**ASSESSMENT PERFORMANCE**

85% Average Score (38.2/45)

29% Low Score (13/45)

100% High Score (45/45)

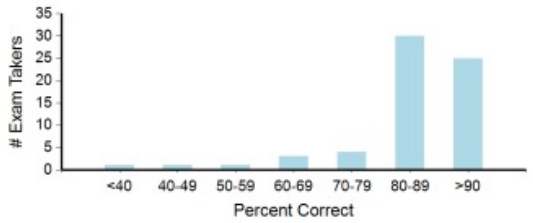
**Assessment Score Reliability (KR-20)**

0.0 0.83 1.0

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2019

**ASSESSMENT PERFORMANCE**

82% Average Score (31.2/38)

24% Low Score (9.2/38)

106% High Score (40.2/38)

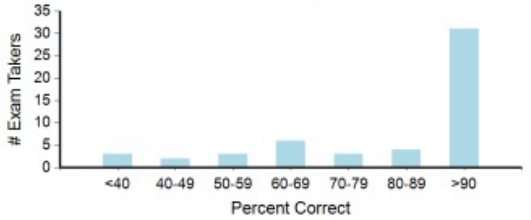
**Assessment Score Reliability (KR-20)**

0.0 0.91

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2020

**ASSESSMENT PERFORMANCE**

77% Average Score (31.5/41)

7% Low Score (3/41)

100% High Score (41/41)

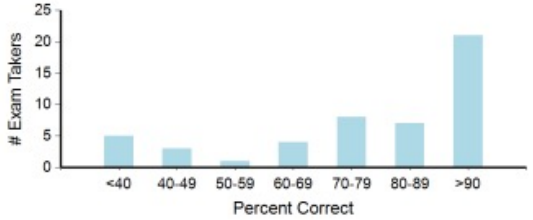
**Assessment Score Reliability (KR-20)**

0.0 0.93

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2021

**ASSESSMENT PERFORMANCE**

85% Average Score (21.3/25)

36% Low Score (9/25)

100% High Score (25/25)

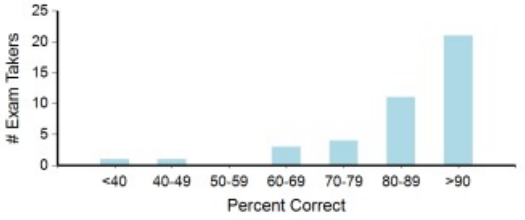
**Assessment Score Reliability (KR-20)**

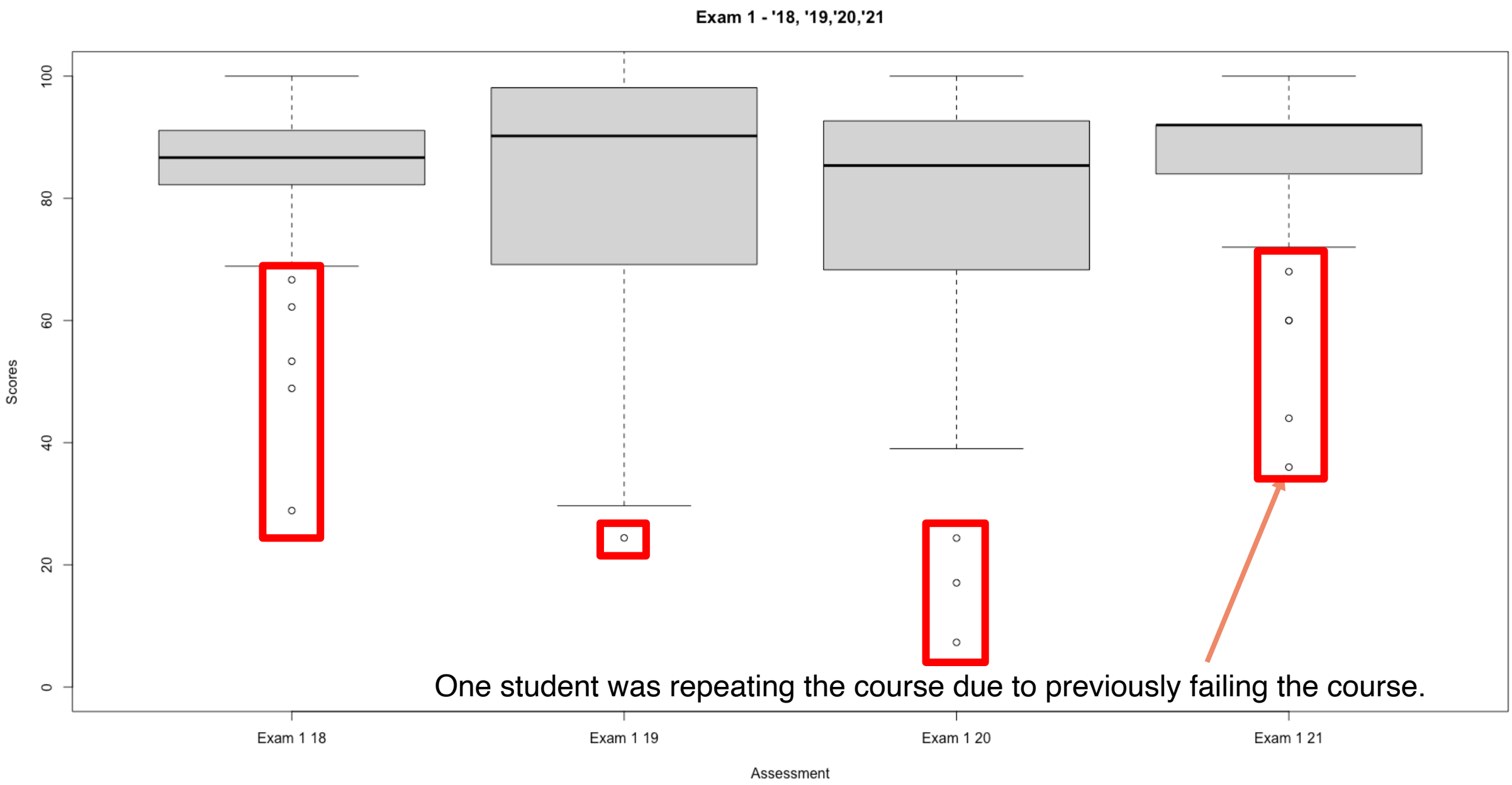
0.0 0.81 1.0

POOR SATISFACTORY GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




Exam 1 - '18, '19,'20,'21
Scores
0
20
40
60
80
100
Exam 1 18
Exam 1 19
Exam 1 20
Exam 1 21
Assessment
One student was repeating the course due to previously failing the course.

# Exam 2 Overview – 2018-2021

## 2018

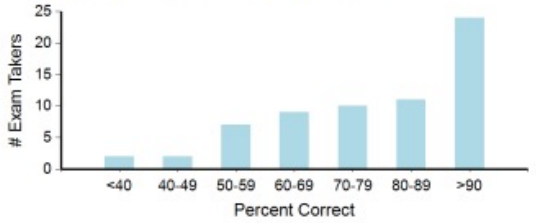
**ASSESSMENT PERFORMANCE**

77% Average Score (30.9/40)
25% Low Score (10/40)
100% High Score (40/40)

**Assessment Score Reliability (KR-20)**

0.0 0.89 1.0
POOR SATISFACTORY GOOD
*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2019

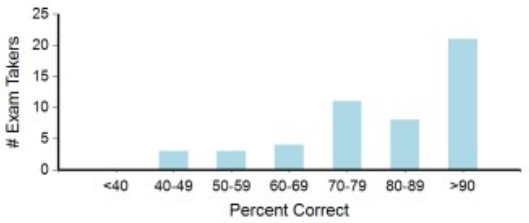
**ASSESSMENT PERFORMANCE**

83% Average Score (29.1/35)
47% Low Score (16.4/35)
107% High Score (37.4/35)

**Assessment Score Reliability (KR-20)**

0.0 0.85 1.0
POOR SATISFACTORY GOOD
*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2020

**ASSESSMENT PERFORMANCE**

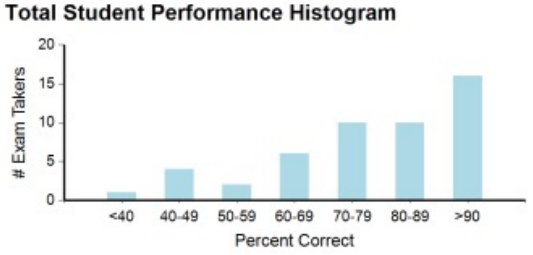
79%* Average Score (26.8/34)*
29%* Low Score (10/34)*
103%* High Score (35/34)*

*Bonus points were available on this assessment and were calculated into the score.

**Assessment Score Reliability (KR-20)**

0.0 0.87 1.0
POOR SATISFACTORY GOOD
*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2021

**ASSESSMENT PERFORMANCE**

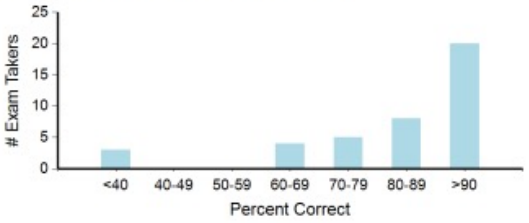
84%* Average Score (20.2/24)*
29%* Low Score (7/24)*
104%* High Score (25/24)*

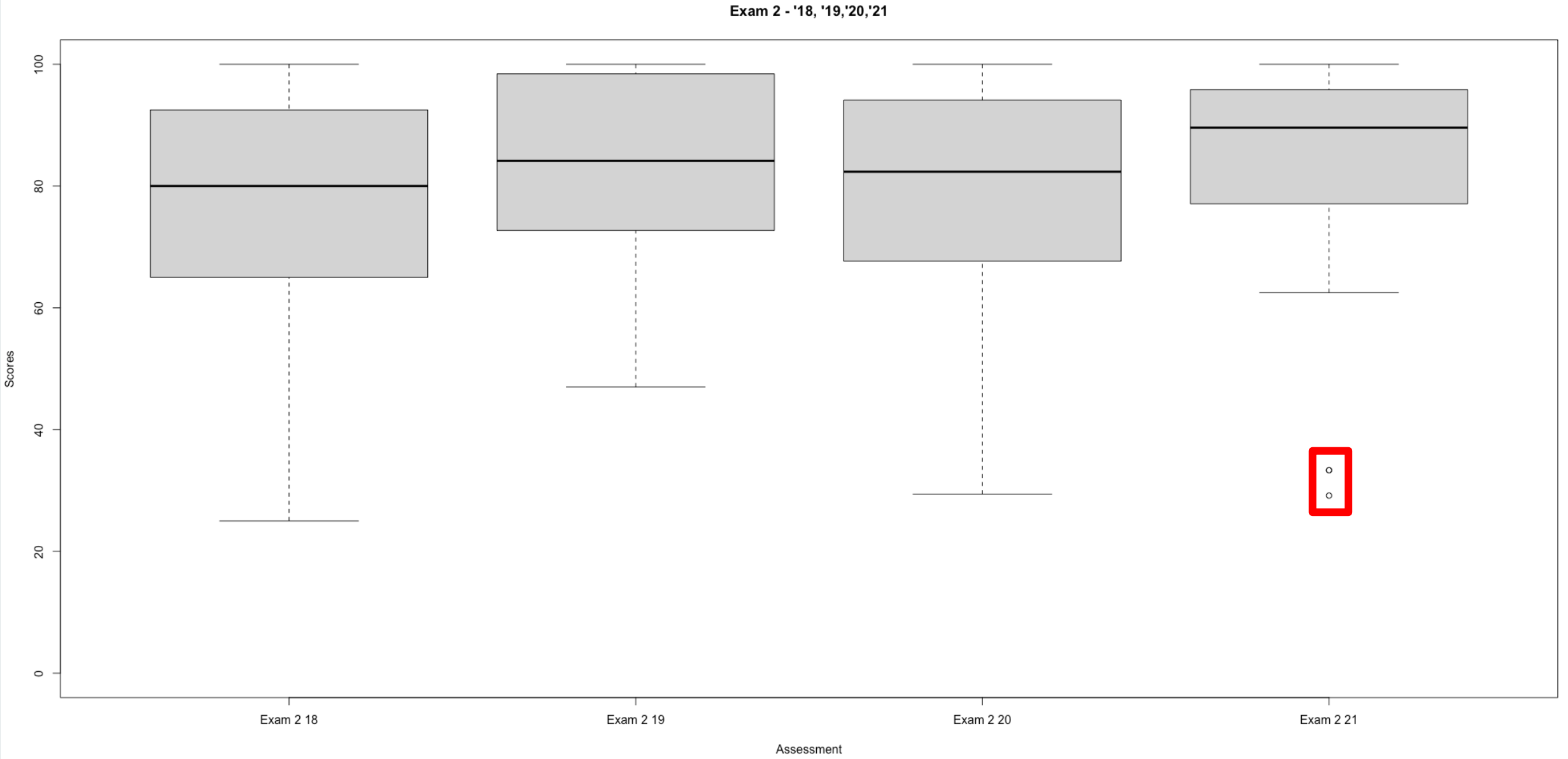
*Bonus points were available on this assessment and were calculated into the score.

**Assessment Score Reliability (KR-20)**

0.0 0.87 1.0
POOR SATISFACTORY GOOD
*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




Exam 2 - '18, '19,'20,'21
Scores
0
20
40
60
80
100
Exam 2 18
Exam 2 19
Exam 2 20
Exam 2 21
Assessment

# Final Exam Overview – 2018-2021

## 2018

### ASSESSMENT PERFORMANCE

80% Average Score (40.1/50)

32% Low Score (16/50)

98% High Score (49/50)

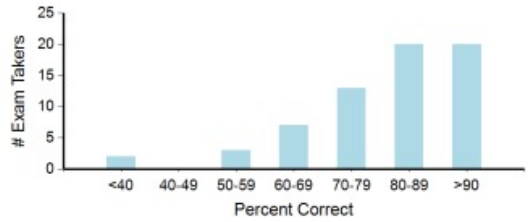
**Assessment Score Reliability (KR-20)**

0.0 — 0.87 — 1.0

POOR | SATISFACTORY | GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2019

### ASSESSMENT PERFORMANCE

83% Average Score (33.2/40)

55% Low Score (22/40)

100% High Score (40/40)

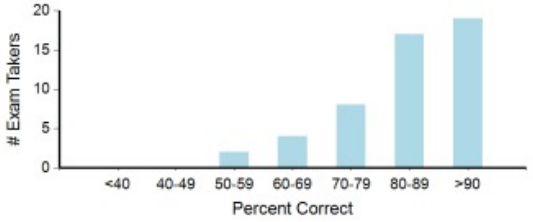
**Assessment Score Reliability (KR-20)**

0.0 — 0.78 — 1.0

POOR | SATISFACTORY | GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2020

### ASSESSMENT PERFORMANCE

81%* Average Score (32.4/40)*

18%* Low Score (7/40)*

103%* High Score (41/40)*

*Bonus points were available on this assessment and were calculated into the score.

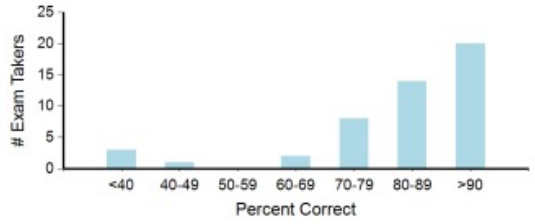
**Assessment Score Reliability (KR-20)**

0.0 — 0.9

POOR | SATISFACTORY | GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




## 2021

### ASSESSMENT PERFORMANCE

92%* Average Score (35.9/39)*

62%* Low Score (24/39)*

103%* High Score (40/39)*

*Bonus points were available on this assessment and were calculated into the score.

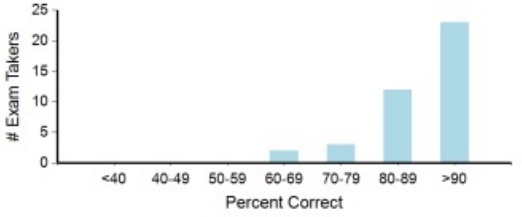
**Assessment Score Reliability (KR-20)**

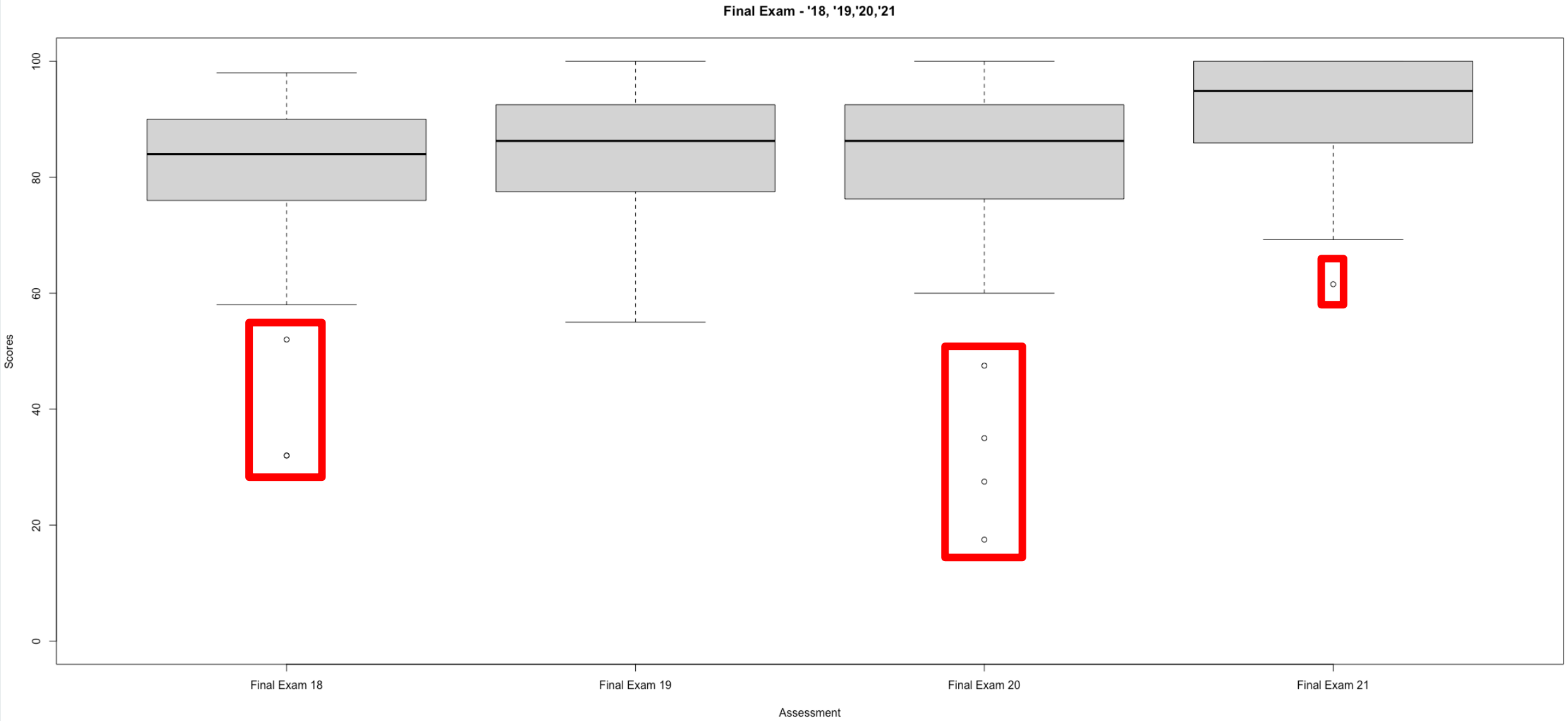
0.0 — 0.83 — 1.0

POOR | SATISFACTORY | GOOD

*Likelihood of students repeating the same performance.*

**Total Student Performance Histogram**




Final Exam - '18, '19,'20,'21
Scores
0
20
40
60
80
100
Final Exam 18
Final Exam 19
Final Exam 20
Final Exam 21
Assessment

# Questions?

# Sources

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. Association for Talent Development.

# Our contact information

Dr. Kimberly Daugherty
Kdaugherty@sullivan.edu

Dr. Sarah Raake
sraake@sullivan.edu

Ben Stephens
bstephens@sullivan.edu