

Generating Effective Feedback from Online Assessments

Okan Bulut

Guher Gorgun, Seyma Nur Yildirim Erbasli, Tarid Wongvorachan

Measurement, Evaluation, and Data Science

University of Alberta



bulut@ualberta.ca



www.okanbulut.com



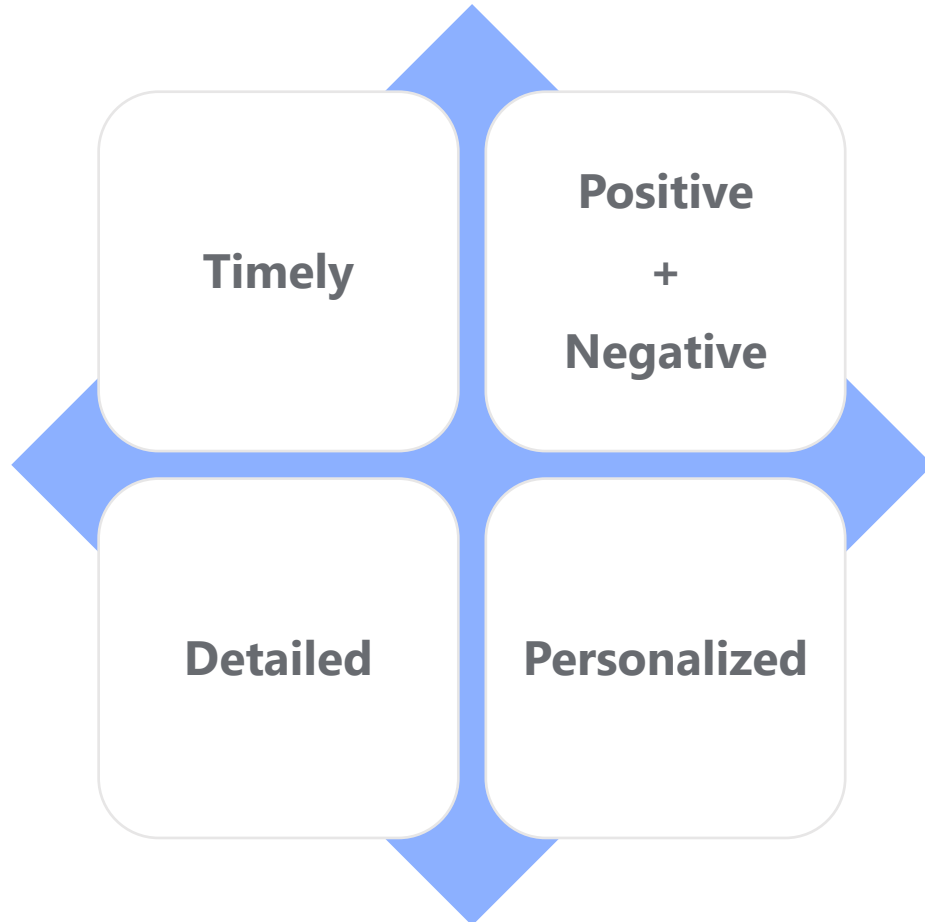
[@drokanbulut](https://twitter.com/drokanbulut)

Why Online Assessments?

- **Diversity in item format**
 - Multiple-choice, True/False, matching, short-answer items, essays, etc.
- **Objective and quick marking**
- **Test security**
 - Randomized questions and response options
- **Additional data on test-taking behavior**
 - Response time
 - Records of response changes



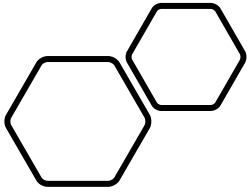
Effective Feedback



Automated feedback generation

- **Exam scores**
 - A total score
 - Subscores for each content area
- **Feedback format**
 - Visual feedback
 - Written feedback
- **Level of detail**
 - Feedback on each content area
 - Key concepts based on each question






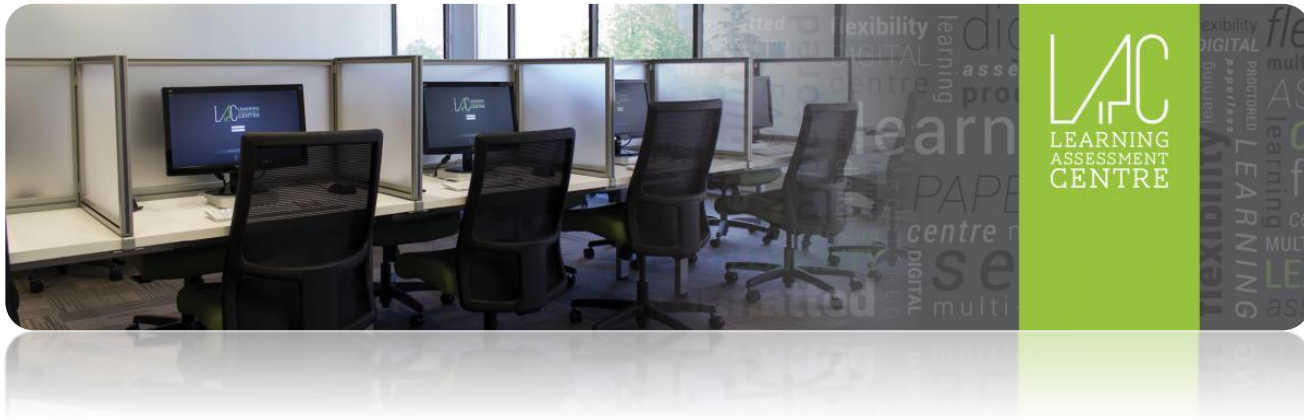
Predecessor of
examvis






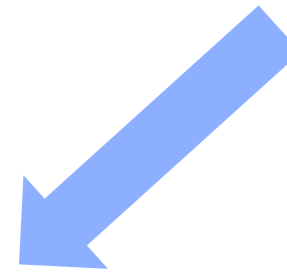
Assessment Blueprint



Question	Content	Key Concept
1	Lecture 2	Difference between LO and IO
2	Lecture 1	Measurement error
3	Lecture 3	Elements of validity
4	Lecture 3	Difference between reliability and validity
5	Lecture 3	Difference between reliability and validity
6	Lecture 3	Difference between reliability and validity
7	Lecture 4	Goals of formative assessment
8	Lecture 4	Differences between formative and summative assessment
9	Lecture 4	Goals of formative assessment
10	Lecture 2	Learner outcomes
11	Lecture 2	Writing appropriate instructional objectives
12	Lecture 4	Test blueprint
13	Lecture 3	Threats against reliability
14	Lecture 2	Bloom's taxonomy
15	Lecture 4	Goals of summative assessment
16	Lecture 3	Elements of validity
17	Lecture 1	Threats against reliability
18	Lecture 1	Measurement error
19	Lecture 3	Elements of validity
20	Lecture 2	Learner outcomes
21	Lecture 2	Goals of instructional objectives
22	Lecture 3	Reliability-validity relationship
23	Lecture 1	Elements of classroom assessment
24	Lecture 2	Bloom's taxonomy
25	Lecture 2	Bloom's taxonomy
26	Lecture 2	Bloom's taxonomy
27	Lecture 4	Preparing summative assessment
28	Lecture 2	Writing appropriate instructional objectives
29	Lecture 4	Implementing formative assessment
30	Lecture 3	Elements of reliability



Edit settings
Group overrides
User overrides
 Edit quiz
 Preview
 Results
<input type="radio"/> Grades
<input checked="" type="radio"/> <u>Responses</u>
<input type="radio"/> Statistics





EDPY 303

Midterm 1 Report
by Dr. Okan Bulut

©2019 - Okan Bulut

Contents

Introduction	3
Classroom Distribution	4
Your Performance by Lectures	5
Your Feedback by Lectures	6
Key Concepts to Review	7

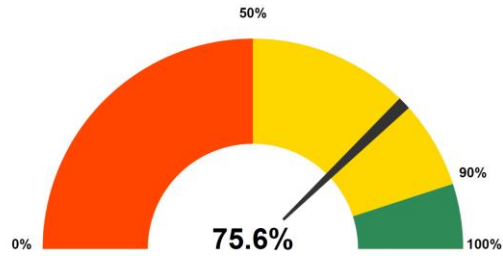
Introduction

This score report aims to help you better understand the results of Midterm 1 and identify the content areas that you should review again before the final exam. The report includes a summary of your overall performance on the exam as well as your performance by lectures to give you a clear picture of in which content areas you have performed better. Additionally, you will find the classroom averages for both the total exam score and the scores by lectures – which will help you view your relative standing in the class. Please review your report carefully and let me know if you have any further questions.

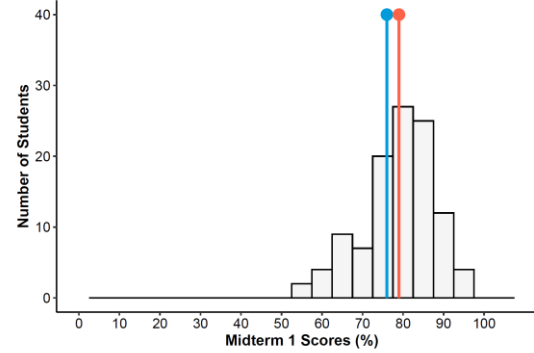
Name: ██████████

Raw Score: 34 out of 45 (Class Average = 36)

Percent Score: 75.6% out of 100% (Class Average = 78.9%)



Classroom Distribution



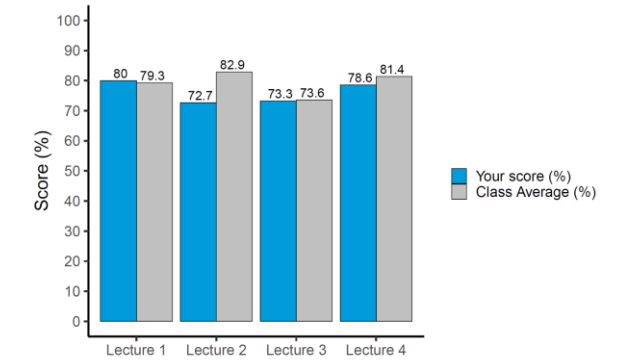
(Note: Blue line = Your score; Red line = Class Average)

The graph above demonstrates the distribution of the total scores on Midterm 1. The height of the gray bars represents the number of students and the x-axis represents the possible scores on the exam (i.e., 0% to 100%). The scores with relatively higher gray bars represent the scores that most of your classmates obtained in Midterm 1. In the graph, the blue line represents your total score on the exam and the red line represents the classroom average. This can help you see if your score is above or below the classroom average. If either the red or blue line is **not** visible, it means that your score and the classroom average are either nearly identical or exactly the same.

On the next page, you will find a table and a bar graph summarizing your percent scores by lectures. Please note that both the table and the graph present **the same information**. In the bar graph, the higher the blue bars, the better your performance on the questions from each lecture. Also, you can compare your scores (blue colour) for each lecture against the average scores by lectures based on the whole class (gray colour).

Your Performance by Lectures

Lectures	Number of Questions	Your Percent Score (%)	Class Average (%)
Lecture 1	5	80.0	79.3
Lecture 2	11	72.7	82.9
Lecture 3	15	73.3	73.6
Lecture 4	14	78.6	81.4



On the following pages, you will find your written feedback based on your performance on the questions from each lecture. In addition, you will see a list of key concepts that you should consider reviewing again before the final exam.

Your Feedback by Lectures

Dear [REDACTED]

Below you will find my comments about your performance on the questions related to each lecture. Please feel free to let me know if you have any questions after you review my comments and the list of key concepts at the end of this report.

Lecture 1: In this lecture, we talked about the primary characteristics of classroom assessments (e.g., purpose, measurement and evaluation, and use) and why we (teachers) need to develop high-quality assessments. The results of Midterm 1 indicate that you did a great job on the questions related to the content covered in Lecture 1. Now, you may start seeking opportunities to use this knowledge as you make your way into schools at the end of this semester.

Lecture 2: Learner outcomes and instructional objectives were the focus of Lecture 2. We talked about learner outcomes as the guidelines provided by Alberta Education and instructional objectives as the statements teachers need to write before designing their instructional activities and assessments. To write appropriate instructional objectives, we need to make sure that they are measurable, observable, and concise. Furthermore, we can benefit from Bloom's taxonomy to target different cognitive levels (e.g., remembering, understanding, and applying). Please note that these cognitive levels represent the cognitive complexity (i.e., type of cognitive skills students have to utilize), not necessarily the difficulty level of assessment tasks. The results of Midterm 1 show that you seem to have missed some key ideas related to learner outcomes, instructional objectives, and Bloom's taxonomy. Before the final exam, you may want to focus more on the key concepts listed below that highlight areas for further review. I want to make sure that you have a solid understanding of learner outcomes and instructional objectives before finishing EDPY 303. Therefore, please do not hesitate to contact me or the teaching assistants should you have any further questions about Lecture 2.

Lecture 3: In Lecture 3, we talked about three important concepts for classroom assessment: reliability (accuracy and consistency of assessments), validity (content of assessment and use of assessment results), and fairness (providing students with equal opportunities to demonstrate their learning). Midterm 1 included questions on the relationship between reliability and validity; internal and external threats against reliability, and ways to collect evidence supporting validity of classroom assessments. The results of Midterm 1 show that the questions about reliability, validity, and fairness were a bit of weak spot for you. This is the case for most students taking EDPY 303 because you have just started to learn about these concepts and understanding what these concepts actually mean will take some time. The next few lectures after Midterm 1 will help you understand these concepts better as we will focus on reliability and

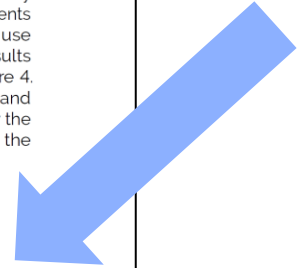
validity in the context of different assessment tools. Before the final exam, I recommend you to revisit the key concepts highlighted in the table below. In addition, you can always contact me or the teaching assistants if you need any clarification regarding these key concepts covered in Lecture 3.

Lecture 4: It is important for teachers to know about the types of classroom assessments, especially those administered during and after instruction. In Lecture 4, we talked about pre-instruction, formative, and summative assessments. Pre-instruction assessments are mostly informal assessments to identify students' levels before starting to teach a new topic/unit (or a new semester); formative assessments are typically ungraded learning activities that the teacher can use during instruction to give students feedback; summative assessments are the real assessments that we (teachers) use for marking and assigning grades after teaching a particular topic or unit. The results of Midterm 1 show that you struggled a bit with the questions regarding Lecture 4. Your understanding of the types of classroom assessments (especially formative and summative assessments) is very important. Therefore, I recommend you to review the key concepts regarding Lecture 4 (listed in the table below) and contact me or the teaching assistants for further clarifications regarding these key concepts.

Key Concepts to Review

The following list shows the key concepts that you should review again before the final exam. The list has been generated based on the questions that you answered incorrectly on Midterm 1.

Lectures	Key Concepts
Lecture 1	Goals of classroom assessment
Lecture 2	Writing appropriate instructional objectives, Goals of instructional objectives
Lecture 3	Elements of validity, Reliability-validity relationship, Importance of validity
Lecture 4	Goals of formative assessment, Formative assessment strategies, Differences between formative and summative assessment



“I just wanted to say thank you for doing this. I have never had a professor who I feel has cared about my individual achievement. Nothing but the utmost respect and appreciation for you.”

“Thank you very much for the great feedback from midterm 1. This is the most useful exam I have ever written and feedback I have reconciled. I appreciate the detailed responses and feedback to help me do better in coming exams and in future application.”

“Thank you so much for this detailed report! I have never gotten feedback for a multiple-choice exam before, and this is greatly appreciated.”

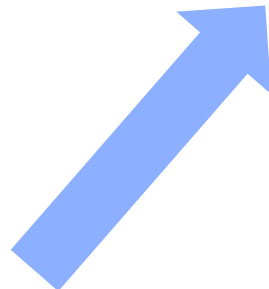


<https://bit.ly/examvisdocumentation>

Setting up **examvis**

1. Add the **examvis** plugin to your eClass course page
2. Create a quiz on eClass
 - Multiple-choice, True/False, open-ended questions
3. Group your exam questions
 - Create **at least 3** content categories (better visualizations!)
 - Each content category should have **at least 3 or more** questions.
 - Move questions into these categories
 - Enter **at least 1** key concept for each question

- ⚙️ Edit settings
- Group overrides
- User overrides
- ⚙️ Edit quiz
- 🔍 Preview
- 📊 Results
- Grades
- Responses
- Statistics
- Manual grading
- Locally assigned roles
- Permissions
- Check permissions
- Filters
- Logs
- Backup
- Restore
- Question bank
- Questions
- Categories**

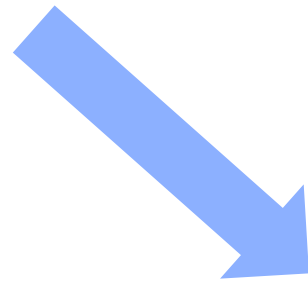


Questions **Categories** Import Export

Edit categories ?

Question categories for 'Quiz: CTT Quiz'

- **Default for CTT Quiz (0)**
The default category for questions shared in context 'CTT Quiz'.
⚙️
- **Lecture 1 (8)** 🗑️ ⚙️ ← ↓
- **Lecture 2 (8)** 🗑️ ⚙️ ← ↑ ↓ →
- **Lecture 3 (15)** 🗑️ ⚙️ ← ↑ →



Create a new question ...

	Question	Actions	Created by
<input type="checkbox"/>	Question name / ID number		First name / Surname / Date
<input type="checkbox"/>	Q11	Internal consistency Edit ▾	Okan Bulut 3 October 2020, 6:20 PM
<input type="checkbox"/>	Q13	Internal consistency Edit ▾	Okan Bulut 3 October 2020, 6:22 PM
<input type="checkbox"/>	Q2	Score interpretation in CTT Edit ▾	Okan Bulut 3 October 2020, 5:51 PM

If a test is very easy for a given group, we would expect internal consistency to be _____.

If you administered a test of English grammar to **all of the students** in a high school (including grades 9, 10, and 11), for w

CTT Quiz

Last Modified: 2/6/2021, 1:01:43 PM

✓ Max Marks: 50

📁 Categories: 3

📄 Total Questions: 31

🔒 Opens At: 2/3/2021, 9:00:00 AM

🔒 Closes At: 2/3/2021, 12:00:00 PM

Edit Feedback

Generate Reports

15 Reports Available

Questions

Feedback Settings

Threshold ≥ 80 %

Message for students at or above the threshold

Your performance in the items related to Lecture 1 was great! Please review the list of key concepts at the end of this

Message for students below the threshold

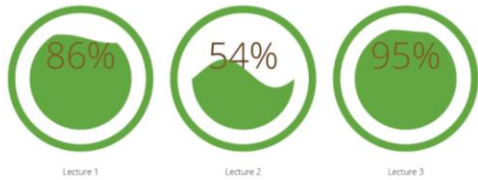
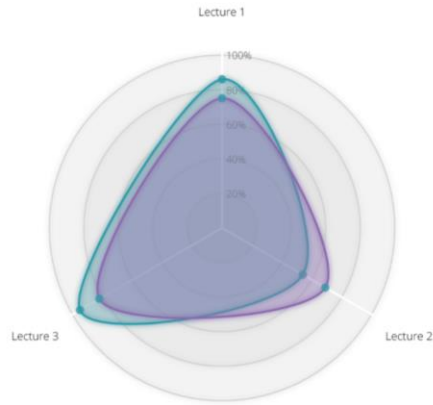
Your performance in the items related to Lecture 1 was not sufficient. You should review the lecture notes as well as t

Save

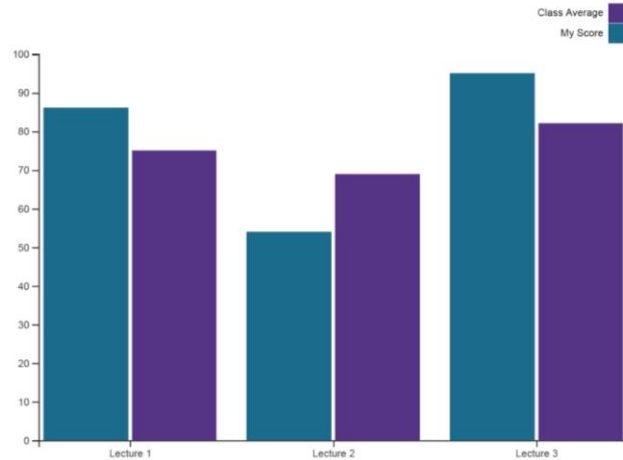


EDPY 507 (LEC B01 Winter 2021) CTT Quiz → 82%
40.75 out of 50

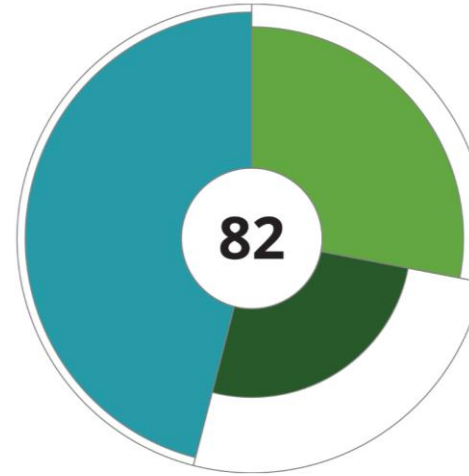
Radar Chart summarizes your performance in the exam based on various content categories (e.g., chapters, lectures, etc.) defined by your instructor. The blue area shows your content-based scores, while the purple area shows the classroom average for the same content categories. The larger the blue area, the better your performance in the exam. The content categories where your score did not reach a high percentage score indicate the content areas that need further study.



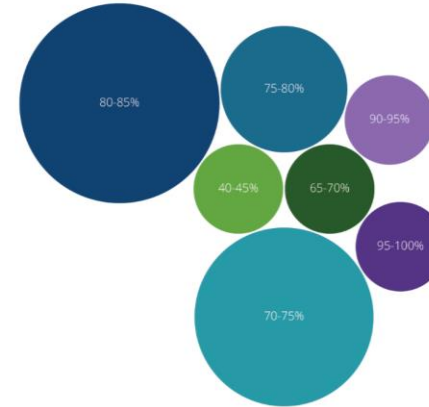
Liquid Fill Gauge shows your scores in the exam based on the content categories defined by your instructor. Each green gauge shows your percentage score in a particular content category. The higher your percentage score is, the more complete the liquid gauge will be. *Liquid Fill Gauge* allows you to compare your performance across the content categories and identify your strengths and the content areas that need further study.



Bar Chart conveys the same information as Radar Chart in a different format. Your percentage scores for the content categories are presented as blue bars, while the classroom average for the same content categories are presented as purple bars. For each content category, you can see your percentage score as well as the classroom average by comparing the height of the bar against the percentage axis. *Bar Chart* provides an overall picture of your performance relative to the classroom average.



Aster Plot (percentage) is presented in the middle of Aster Plot. Each slice above the middle point represents a content category. The width of the slices shows the weight of the content category. The wider the slice, the more questions this content category has in the exam. Content categories with more questions would be more influential on your overall performance. The height of the slices shows your performance in a given content category. The higher the slice, the better your performance in that category. Please move your mouse over the graph to see what the layers represent.



Bubble Chart shows the distribution of overall test scores from all students who attended the exam. Each bubble represents a range of possible scores in the exam (e.g., 100% to 96%; 95% to 91%; 90% to 86%, etc.). The larger the bubble, the more students are located in that particular score range. *Bubble Chart* aims to provide a snapshot of the overall test scores from the exam and enable students to see their positions in the exam relative to other students.

Lecture 1	Your performance in the items related to Lecture 1 was great! Please review the list of key concepts at the end of this report. These concepts refer to individual items that you have missed on the quiz.	In this section, you will see detailed feedback for all of your major content categories for this exam.
Lecture 2	Your performance in the items focusing in Lecture 2 was less than 80%. I recommend you to review the lecture notes as well as the list of key concepts at the end of this report before the final exam. Please feel free to let me know if you have any questions.	
Lecture 3	Congratulations! Your percent-correct score on the items related to Lecture 3 was higher than 85%.	
Lecture 1	Norm and criterion-referenced tests	In this section is a list of some key concepts you may wish to address, along with the content category in which they occurred on the exam.
Lecture 2	Score interpretation in CTT, Internal consistency	
Lecture 3	Item analysis, Discrimination, Coefficient alpha, SEM, Spearman-Brown formula	

What did students think about **examvis**?

- Fall 2020
 - EDPY 303 – Educational Assessment (2 sections)
 - EDPY 507 – Measurement Theory I
- Winter 2021
 - EDPY 303 – Educational Assessment (4 sections)
 - EDPY 507 – Measurement Theory I
 - NURS 125, 216, and 224

What did students think about **examvis**?

- Did you find the information provided in the score reports helpful to better understand your exam performance?
 - **95%-97%**
- Are you planning to use the information presented in your score reports to prepare for your subsequent (or final) exams?
 - **75%-94%**
- Overall, do you think ExamVis has improved your learning in the course?
 - **75%-80%**

What's next?

We are currently investigating (using log data on eClass):

1. How frequently students accessed **examvis**
2. Whether **examvis** affected students' course activities
3. Whether **examvis** improved students' course performance

We will also prepare additional materials on how to design better online assessments and share them as instructional modules.



Thank You!

For questions/comments:

bulut@ualberta.ca