

Project 3: Data-driven improvements

Due Apr 28 by 11:59pm **Points** 100

Learning objectives: Students should be able to

LO1 Apply crowdsourcing or other participatory techniques to improve learning materials and resources

LO2 Apply principles of data-driven improvement to create better learning materials over time

LO3 Articulate requirements for creating an explicit learner model for data driven improvement

Objectives and Background

Developing good quizzes takes time and knowledge about common student misconceptions. Both are limited. In this assignment, you'll create a better quiz, one that is directly based on students revealed knowledge misconceptions, and one that students can take multiple times without much additional effort from either instructors or learners. The secret to this free lunch? Using data that is already being generated in everyday classroom activity.

Why this project goal: This project is an opportunity to integrate all that you've learned about tools for learning. Doing this project successfully requires skills in design, learning science, and technology -- so few other individuals know these as well as you, making this a great project to put on your portfolio!

Note: This is not a project you can do last minute. To succeed in this project, you need to start early.

What you'll do

This assignment takes as raw material quizzes that you've done in this class. In particular, it uses the Question-Response-then-Choose questions you have encountered throughout the semester. These are questions which ask you an open-ended question (Question), and ask you to write a free-text response (Response), followed by a number of choices, where you are asked to pick the choice closest to your free-text response (Choose).

Using data from these questions as the starting point, you'll create new multiple choice or select all that apply questions. The candidate choices for these questions will be previous student Responses (as defined above). **Note that you can switch a given question from multiple choice to select all that apply and vice versa.** That will be particularly useful if you couldn't identify enough correct/incorrect choices from the given

data.

Ideally, we would like to create quizzes where:

1. Students get feedback about whether they are right or wrong immediately after answering a question. Where possible, it would be useful to also present any corrective feedback.
2. Students are allowed to attempt the question multiple times. If they get it wrong initially the options for subsequent attempts would preferably be different each time.

Here's an example from the Deliberate Practice week:

Q:

We are building an iPad app to teach factory workers to quickly and accurately identify specific defects on a circuit board(e.g. dents, chips etc.) with a visual inspection. Our goal is a high level of performance (accuracy and speed). Describe a suitable app design to do this in three sentences.

Here is an example Response to this Question.

Show a defective question, get the answer, give feedback about why correctness and reason, then repeat.

We would like to use such Responses to construct options for a multiple choice question.

For example, using the Responses to this question, here are four options:

1. *Show a defective question, get the answer, give feedback about why correctness and reason, then repeat.*
- 1. *Provide explicit instruction about the best method; 2. Supervised by tutors to diagnose errors; 3. Provide informative feedback and remedial training.*
 - Noticing defects is a matter of noticing subtleties between what's functional and what's not. Contrasting cases, along with immediate feedback over multiple example would do an excellent job at teaching the difference.
 - By providing contrasting cases, workers will be able to spot differences by comparison. Showing the right answer is necessary for learners to accurately identify the defects.

Note that all four options are directly taken from previous Responses as written by students. Because these options come from what students wrote as answers, they reflect student understanding of the concepts

being tested. Note also that this example isn't perfect -- there are two options talk about contrasting cases for instance, and it's not clear how they are different.

Data

First, is data about the questions. This file has the following fields:

Q_id A unique id for the question.

Q_text The actual text of the question.

Answer_choice What is the correct answer for the choose step? E.g. "A", or "A,B"

Choice_A: Text for Choice A

...

Choice_D: Text for Choice D

Second, there is data about the answers. This file has the following fields:

Q_id The id of the question this is the answer for

Score What did the student score for the choose answer step? Between 0 and 1. 1 means fully correct answer, numbers in between represent partial credit.

Score_overall What did this student score on this quiz as a whole? Between 0-10

Answer_text Free response answer to the question

Answer_choice Chosen answer for choose step. E.g. "A", or "A,B"

We will provide you with these two files in two sets. The first is a "development" set. Use this set to develop your scheme for generating multiple choice questions.

The second set is the "validation" set. Run your code you created with the development set on this set. Note what worked well and what did not. You will need to include this report in your project submission.

To discourage you from peeking at the validation set while developing your code (and to make it easier on us) we will provide this validation set the week of Apr 15.

How to develop quizzes

The data that we've given you can help you develop simple rules (heuristics) about which student free response answers we could reuse as options. For instance, you might decide to pick as your "correct" option the response from a student who got a high score on this question in the quiz. Similarly, you may choose as distractor options responses from students who got no credit for their "choose" answer to this question.

The precise method you use is up to you. There are only two requirements:

- You should provide the code that you wrote for the method. (So, you can't pick options manually.)
- Your method should not be specific to quizzes from our class. Given a set of files with the right format, you should be able to develop quizzes for any course.

You can use any programming language to do this data driven work. In the past, students have enjoyed working with Python and Jupyter notebooks.

Deliverables:

There are four deliverables:

1. A link to a working prototype webpage. Webpage can display only one question at a time, or all at once. But it should show feedback immediately after students attempt each question. We will test your prototype on Firefox 65 or later on Mac OS.
2. Code for generating questions. Please host this on Github or similar site, or on a hosted Jupyter notebook (e.g. <https://colab.research.google.com/notebooks/welcome.ipynb> (<https://colab.research.google.com/notebooks/welcome.ipynb>)).
3. Write-up 1, we ask you to summarize your method for:
 - Which response to pick for "correct" option
 - Which response to pick for incorrect options
 - How do you pick options for students who get the question wrong initially? (Retry options)
 - What feedback do you show immediately after students attempt a question.

For each choice, justify your choice in terms of learning science theory, or other relevant concepts you have learned in this/related classes. If you're using concepts from a different class or other research, please include a pointer to the original material.

4. Write-up 2: Take the code you wrote using the development set, and run it on the validation data. Note what worked well and what did not. Do not make changes to your code at this stage; but add in your write up: what would you do to improve your method further.

In summary, Write-up 2 talks about: what worked well and what did not, and what you might do to improve further (this could go beyond what you observed didn't work well with the validation set.)

Don't worry about

In your project development, don't worry about:

- Formatting/grammar issues with student Responses
- Ordering options in multiple choice
- Being able to only support a limited number (>1) of retries.
- Working on screens smaller than 11 inches

FAQ

1. May I change teams?

Yes.

2. What does "development set" and "validation set" mean?

The development set is the data you use to build your model and train it. After you are done and your model is ready (e.g. selecting correct options), you want to test it. That is when you use a different set of data that you didn't train your model on which is the validation set. So the validation set is the data you test your model on to make sure it's not only functional on the development set and you can use on other data.

Note: as we get more questions from you we will update the FAQ with answers and more tips. Please check back frequently.

Data-driven project					
	%	Criteria/ Ratings	Low Mastery 1	Medium Mastery 3	High Mastery 5
	Prototype				
	15%	Technical functionality: 1- Select choices both correct and	The prototype choice selection was manual or random and doesn't make a distinction between correct and incorrect	The prototype is following an automatic mechanism of selection for either correct or incorrect choices but not both	The prototype is following an automatic mechanism of selecting correct and incorrect choices for the different questions

50%		incorrect choices			
15%	Technical functionality: 2- Ability to retake a question at least once	The prototype doesn't allow for a retake of questions or doesn't change the options for the retake	The prototype allows for at least one retake of each question but changes are minimal (e.g. shuffle choices)	The prototype allows for at least one retake of each question with one or more options changing	
10%	Technical functionality: 3- Provide feedback after each question submission	The prototype doesn't provide any feedback or provides irrelevant feedback (e.g. that a question was attempted)	The prototype provides relevant feedback after submission of a group of questions but not after each question	The prototype provides relevant feedback for submission after each question	
10%	Generalization of method beyond given data	The prototype only works for questions/answers in the development data set	The prototype generalizes to questions/answers in both the development and validation set	The prototype generalizes and works with any other data that has the same format as the development/ validation set	
Writeup					
10%	Writeup of the choice selection mechanism	The writeup is absent or vague in explaining the mechanism for correct and incorrect choices selection	The writeup explains the prototype selection mechanism for correct or incorrect choices but not both and it vaguely explains the reasons for choosing the mechanism	The writeup explains the prototype selection mechanism for correct and incorrect choices and clearly explains the reasons for choosing the mechanism	
10%	Writeup of the question retake mechanism	The writeup is absent or vague in explaining the question retake mechanism	The writeup explains the question retake mechanism but doesn't provide reasons for choosing it	The writeup explains the question retake mechanism and provides reasons for choosing it	
	Writeup of	The writeup is absent		The writeup explains the	

50%	10%	feedback mechanism supported by theory	or vague in explaining the feedback mechanism and does not show how it apply learning theory	The writeup explains the feedback mechanism in the prototype but does not show how it apply learning theory	feedback mechanism in the prototype and shows how it apply learning theory
	10%	Summary of what worked and what didn't, using the model on the validation set	The summary is absent or vague in describing the outcomes analysis of what worked and what did not work	The summary describes the outcomes analysis of what worked and what did not work without prediction of the possible reasons behind it	The summary describes the outcomes analysis of what worked and what did not work with a prediction of the possible reasons behind it
	10%	Future improvement reflection	The reflection is absent or does not indicate future improvements	The reflection includes future improvements but does not include the anticipated outcomes if they are to be implemented	The reflection includes future improvements and anticipated outcomes if they are to be implemented

you can get reflection, cannot manually add that part, you cannot add new data.
 shouldn't be changing the code
 the feedback doesn't have to be complex.