Quantitative Reasoning Assessment Pilot Project (Fall 2016)

Jessica Chapman and Ivan Ramler (Statistics)

Introduction

The new general education requirements, taking effect in Fall 2013, include a course designated as "Quantitative or Logical Reasoning" (QLR). Courses can be designated as "QLR" by meeting the learning goals for either Quantitative Reasoning or Logical Reasoning. In particular, the new general education requirements specify that:

> 1. Quantitative Reasoning Courses have primary learning goals in which students, through multiple opportunities and classroom instruction, develop their abilities to:
>   a. address questions by examining quantitative evidence using appropriate methods of analysis and evaluation; and
>   b. explain their conclusions and the quantitative methods they used in developing their reasoning.

In the fall semester of 2016, three faculty members from the Department of Mathematics, Computer Science, and Statistics participated in a pilot assessment of the University's QLR general education requirement. Stat 113 (Applied Statistics) was chosen for this pilot project because of its high popularity. The course fulfills St. Lawrence University's new general education requirement for Quantitative Reasoning. In addition, Stat 113 is required for the following majors: Biology, Business in the Liberal Arts, Economics, Neuroscience, Psychology, and Statistics. Other majors, such as Government, increasingly encourage their students to take a course in applied statistics. Of the most recent graduating class (Class of 2017), 68% of all students took this course as part of their program of study.

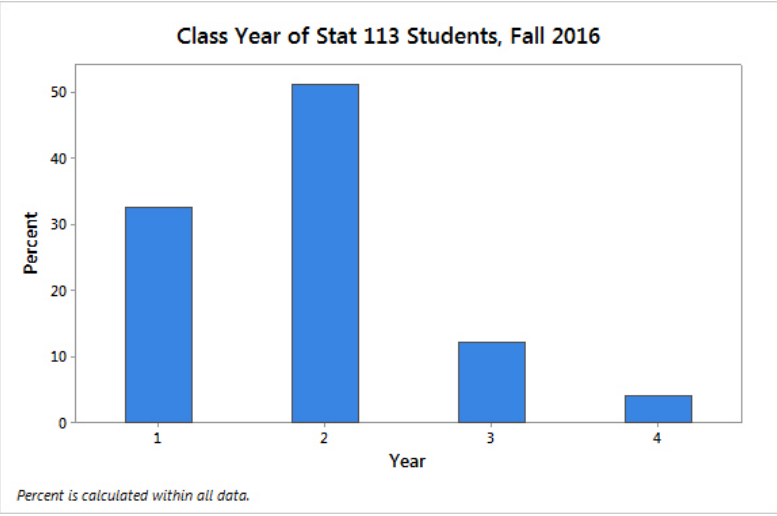

Figure 1: Bar chart of class year for students enrolled in Stat 113 at the beginning of the Fall 2016 semester

As seen in Figure 1, Stat 113 is predominantly taken by first and second year students (over 80% in Fall 2016); this is unsurprising, as it serves as a pre-requisite for courses in Economics, Psychology, and Statistics. Generally, there are more females than males in Stat 113; for the Fall 2016 semester, 60% of students enrolled identified female and 40% identified as male.

Methodology

The three participating faculty members (across four sections of Stat 113) each randomly chose 15 final exams to assess for this pilot project, for a total of 60 students. Final exams were chosen because they are comprehensive; in-class exams or student projects are often more narrow in focus and coverage. Thus, using final exams provides a more thorough means of assessing students ability to identify, and justify, which analyses are most appropriate (the essence of the Quantitative Reasoning learning goals). While the participating faculty were not required to have exactly the same questions on their final exams, they were asked to include questions on their exam that met specific criteria (see Appendix). All participating faculty members scored those items using the same rubric (Appendix).

The rubric used for this assessment project (Appendix) identifies 11 specific learning goals that can be mapped to one of the two primary learning goals outlined in the definition of Quantitative Reasoning courses. Table 1 maps the assessment learning goals to the QR learning goals. The rationale behind these mappings is explained in greater detail in the assessment rubric (Appendix).

Table 1: Mapping of assessment rubric items to the learning goals for Quantitative Reasoning courses

QR Learning Goals	**Rubric Items**
1a) address questions by examining quantitative evidence using appropriate methods of analysis and evaluation	LG1 (What graph?) LG2 (What test?) LG6 (Hypotheses) LG7 (Conditions) LG8 (Appropriate support) LG9 (Make an appropriate decision) LG10 (Make an appropriate conclusion)
1b) explain their conclusions and the quantitative methods they used in developing their reasoning	LG3 (Explain graph choice) LG4 (Appropriate numerical summary) LG5 (Make an appropriate conclusion) LG6-LG11

To assess LG1 and LG2, faculty were asked to include a series of questions on their final exam that provided a description of data to be analyzed (without providing any actual data or summaries), and students were tasked with identifying either the appropriate graphical displays of the data or appropriate inferential statistical methods. As an example, a description provided to students might be “Compare cumulative GPAs for male and female students in the class,” and the appropriate graphical display would be side-by-side boxplots and the appropriate inferential statistical methods would be a two-sample t-test (comparing a numerical variable for two groups of a categorical variable). Because these are short answer questions, participating faculty were able to include multiple descriptions for students to identify. Further, using only a single question (description) for students to identify was deemed insufficient for assessing these learning goals,

as a student could get the correct answer simply by guessing; it is highly unlikely that students would get every question/description correct by guessing alone. Students were rated as "Satisfactory" (2), "Needs Improvement" (1), or "Unsatisfactory" (0) depending upon the percentage of these questions that they answered successfully.

The item for LG3 was designed to assess if students could explain/justify how they knew how to solve the questions described above (for assessing LG1 and LG2). However, in this pilot assessment project, LG3 is omitted as two of the participating instructors did not directly assess the goal on their final exam.

To assess LG 4, faculty were asked to include a question on their final exam that asked students to explain which numerical summaries were most appropriate for a set of numerical data. The goal is for students to recognize that the most appropriate numerical summaries depend on the shape of the dataset and presence/absence of outliers. For example, it could be misleading to use the mean to summarize a dataset of salaries which are typically right skewed (or have high outliers); the mean would be inflated by those unusually high values, and a median would be more appropriate.

We recognize that students may be able to make appropriate conclusions from statistical analyses without being able to carry out those analyses themselves. As such, to assess LG5, faculty members were asked to include a question that provided students with output from a statistical procedure and ask the students to make a meaningful and appropriate conclusion from that output.

As indicated in Table 1, LG6 – LG10 map to both of the QR learning goals, and the most direct way to assess those learning goals is to ask students to carry out a significance test in its entirety. Thus, the participating faculty members use one of the hypothesis tests from their final exam to assess these goals. LG11 is also assessed by looking at a specific part of the students' solution to the hypothesis test – the conclusion.

Results

Figure 2 summarizes the results from the pilot assessment project. In all cases, a score of "2" is considered "Satisfactory" and "0" is considered "Unsatisfactory." In general, across all learning goals students in the sample are most commonly viewed as satisfactory (blue bars in Figure 2).

We find that students perform generally well (more than 2/3 rated as satisfactory) on assessment items related to LG7 – LG9 and LG11, though with LG7 students either address this portion of the hypothesis test either completely or not at all (responses almost all either 0 or 2). Overall, these findings are not surprising. The assessment item related to these items is generally specific or guided – something along the lines of "Conduct a hypothesis to …". Thus, it should be pretty clear to students what exactly they are supposed to do on this question. The slight bimodal nature of the scores on LG7 are indicative of the fact that students either remember to check their conditions (or justify their choice of test) or not. That said, it should be noted that half of the department's faculty teach statistics via simulation methods that do not require the strict assumptions of the traditional parametric statistics tests; that could potentially be reflected in the

scores on this item. LG6, which also pertains to the same assessment item, has the lowest number of satisfactory ratings; this suggests that stating hypotheses could be the weak point of most students' hypothesis tests. We note that LG10 is an aggregate of items LG6 – LG9 and does not provide any additional insight.

Students also do generally well on the assessment items related to LG1, with more than 2/3 of students being rated as satisfactory. The interesting finding here is how this is at odds with the scores for LG2 (only 50% rated as satisfactory) despite the similarity in the assessment items (students are asked to identify a graph for the assessment items for LG1 and an inferential procedure for the assessment items for LG2). Thus, it is possible that some students do not see the connection between graphical displays of data and inferential methods.

LG4 has the most unsatisfactory scores (aside from LG7, mentioned above). This is another case of a more "open-ended" question, and a sizeable number of students (almost 1/3) seem to struggle to recognize when some numerical summaries are better than others. It is also possible that some students don't remember how to approach this type of question, since it is typically covered at the beginning of the semester.

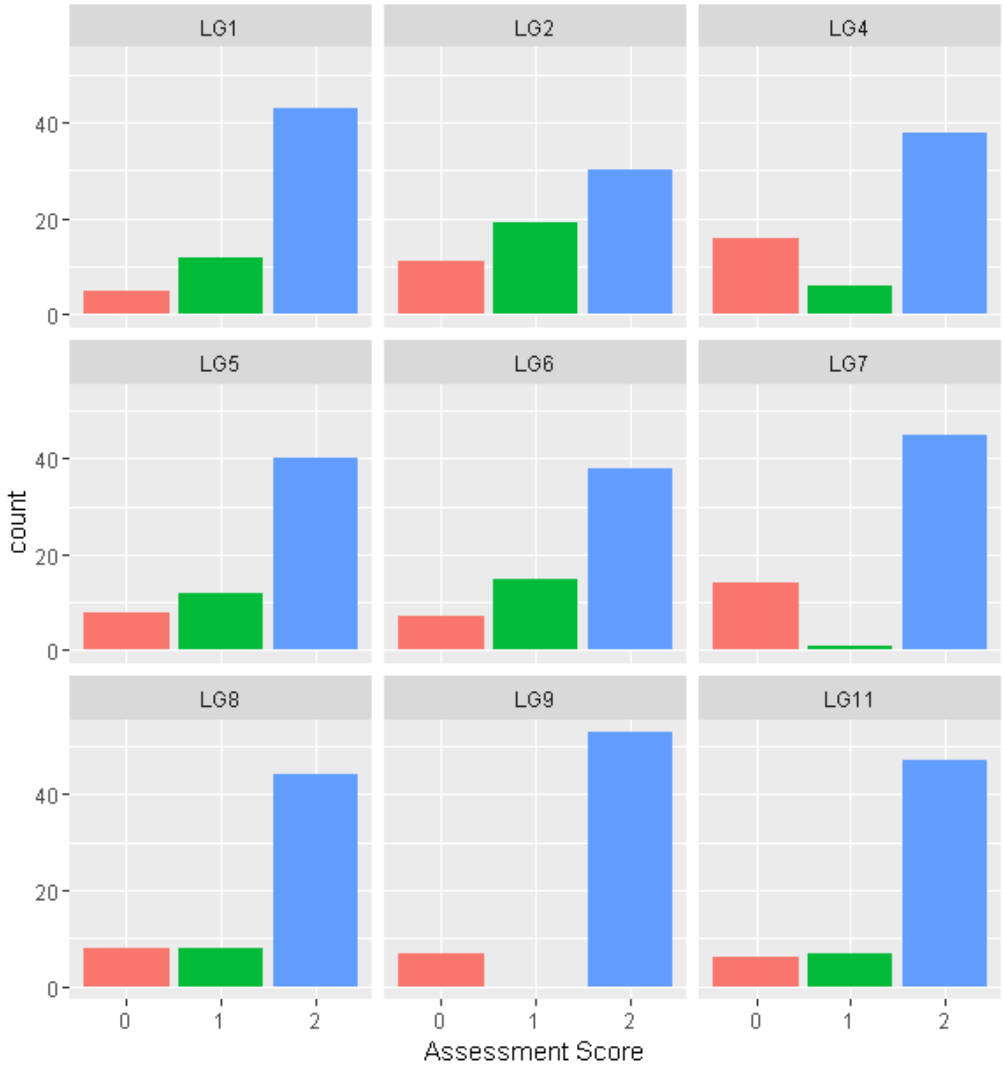

Figure 2: Bar charts of scores across the nine assessed learning goals for Stat 113 and the QLR. Values of 2 indicate a Satisfactory score, 1 – need improvement, 0 – unacceptable.

The slight difference between scores on the assessment items for LG5 and LG11 is a little surprising, given that they ask students to do exactly the same thing (make a meaningful conclusion from a significance test). The difference is that for LG5 all of the output from a statistical test is provided whereas they need to obtain that information for themselves to for LG11. While we should recognize that the difference in scores in the two items is not likely to be statistically significant, it is possible that it reflects that some students struggle reading output from statistical software.

Conclusions

Overall, our students are tending to meet the learning goals of the course (and QLR). We find that students tend to perform best on the most directed assessment items (“Conduct a hypothesis

test to do…") and perform slightly worse on the more "big picture" concepts ("What should you do in this situation?".

Moving forward, first and foremost, we need to conduct these assessments across more Stat 113 instructors and collect data from additional semesters to determine if these observed trends hold across more students. While more data is necessary before drastic course changes are recommended, instructors could start to emphasize (more than they already do!) the big picture ideas across the entire semester. Second, the Statistics faculty should meet (over the summer or early in the fall semester) to review these findings, discuss the current rubric, and make any changes or additions to the rubric items.

Appendix

Draft QLR Rubric (+ Rationale) for Stat 113 Final Exams

From the QLR Description:

1 . Quantitative Reasoning Courses have primary learning goals in which students, through multiple opportunities and classroom instruction, develop their abilities to:

a. address questions by examining quantitative evidence using appropriate methods of analysis and evaluation; and

b. explain their conclusions and the quantitative methods they used in developing their reasoning.

Using Appropriate Methods of Analysis: "Identifying Inference"

The first step to using "appropriate methods" is to identify from the clues available in the description of the data to choose the method of analysis. "Identifying Inference" questions can be used to assess this aspect of QLR.

QLR Learning Goal	Satisfactory (2)	Needs Improvement (1)	Unsatisfactory (0)
LG1: Using Appropriate Methods of Analysis ("Identifying Inference")	80% or more of these questions are answered correctly	Between 60% and 80% of these questions are answered correctly	Fewer than 60% of these questions are answered correctly

Using Appropriate Methods of Analysis and Explaining the Choice of Methods used in Developing Reasoning: "Which Graph Should You Use?" + Follow-up Explanation

The first step to using "appropriate methods" is to identify from the clues available in the description of the data to choose the appropriate graphical displays for data. An appropriate "explanation of the methods used" would reference the clues to watch for (i.e., number of variables and type(s) of variables being used). "What Graph Should You Use?" questions, and a

follow-up question asking students to explain their reasoning, could be used to assess these aspects of QLR.

QLR Learning Goal	Satisfactory (2)	Needs Improvement (1)	Unsatisfactory (0)
LG2: Using Appropriate Methods of Analysis ("What Graph?")	80% or more of these questions are answered correctly	Between 60% and 80% of these questions are answered correctly	Fewer than 60% of these questions are answered correctly
LG3: Explain the Choice of Methods used in Developing Reasoning ("What Graph?")	Explanation includes looking for <u>both</u> 1) the number of variables being used <u>and</u> 2) the type(s) of variables being used	Explanation includes looking for <u>either</u> 1) the number of variables being used <u>or</u> 2) the type(s) of variables being used	Explanation <u>does not address</u> the number of variables being used and the type(s) of variables being used

Explaining the Choice of Methods used in Developing Reasoning: "Which Numerical Summaries?" + Explanation

Different numerical summaries will be appropriate in different situations. Which numerical summary is most appropriate depends on the shape of the distribution and presence/absence of outliers in the data. A question asking which numerical summaries are most appropriate for a given distribution, with a brief explanation, could be used to assess this aspect of QLR.

QLR Learning Goal	Satisfactory (2)	Needs Improvement (1)	Unsatisfactory (0)
LG4: Explain the Choice of Methods used in Developing Reasoning ("Which Numerical Summaries?")	<u>Both</u> correct numerical summaries are chosen <u>and</u> explanation correctly mentions the shape of the distribution and/or presence/absence of outliers	<u>Either</u> the correct summaries are chosen <u>or</u> explanation correctly mentions the shape of the distribution and/or presence/absence of outliers	The numerical summaries are incorrect <u>and</u> explanation does not address the shape of the distribution and/or presence/absence of outliers

Make/explain appropriate conclusions: **Partial Inference Question with Output Provided**

It is possible for students to be able to correctly interpret/explain statistical results, even if they cannot carry out the details of the statistical methods themselves. The ability to correctly make/explain a conclusion could be assess by providing students with output and asking them to make a conclusion, with a justification of their reasoning.

QLR Learning Goal	Satisfactory (2)	Needs Improvement (1)	Unsatisfactory (0)
LG5: Make/explain appropriate conclusions	Correct conclusion is made and the conclusion is correctly justified by referencing appropriate portions of the output	Correct conclusion is made, but justification is weak	Either conclusion is incorrect or no justification is provided.

Address questions by examining quantitative evidence using appropriate methods of analysis and evaluation AND explain conclusions: Hypothesis Testing

Hypothesis tests are one type of questions where students are asked to choose and carry out an appropriate method of analysis and make an appropriate, meaningful conclusion. A hypothesis test can be used to assess these aspects of QLR.

QLR Learning Goal	Satisfactory (2)	Needs Improvement (1)	Unsatisfactory (0)
LG6: --- (Hypotheses)	Hypotheses are clearly and correctly stated with all notation defined	Hypotheses are technically correct, but either not clearly stated or notation is undefined	Hypotheses are incorrect or not stated at all
LG7: --- (Choice/justification of Appropriate Method)	Correct method is chosen and is justified by checking/verifying the appropriate conditions for that method	Correct method is chosen but is not justified by checking/verifying the appropriate conditions for that method	Incorrect method is chosen
LG8: --- (Appropriate numerical support is provided)	Correct test statistic, p-value, and degrees of freedom (if appropriate) are provided	Either some numerical support is missing (i.e., test stat or p-value) or one piece of numerical support is incorrect	No numerical support is provided or all numerical support is incorrect
LG9: --- (Decision about Null Hypothesis)	P-value (or test statistic) is correctly used to make a decision (Reject or not) about the null hypothesis using either the "scale of p-values" or a significance level; if significance level is used, it is clearly stated		Either no decision or an incorrect decision is made about the null hypothesis
LG10: Address questions by examining quantitative evidence using appropriate methods of analysis and evaluation	*Score for this item is the sum of the subparts assessed above; Satisfactory=8*	*Score for this item is the sum of the subparts assessed above; Needs Improvement = 4 – 7*	*Score for this item is the sum of the subparts assessed above; Unsatisfactory < 4*
LG11: Make appropriate conclusions	Hypothesis test includes a clear conclusion that is stated clearly in the context of the problem	Hypothesis test includes a clear conclusion but it is not stated in the context of the problem	No conclusion is present or conclusion is incorrect