Alignment Between Learning Objectives and Assessments in a Quantitative Literacy Course

Younggon Bae
Michigan State University, baeyoun3@msu.edu
Samuel L. Tunstall
Michigan State University, stunstal@trinity.edu
Kathryn S. Knowles
Michigan State University, appenze2@msu.edu
Rebecca L. Matz
Michigan State University, matz@msu.edu

Follow this and additional works at: https://scholarcommons.usf.edu/numeracy

Part of the Curriculum and Instruction Commons, Educational Assessment, Evaluation, and Research Commons, Scholarship of Teaching and Learning Commons, and the Science and Mathematics Education Commons

Recommended Citation

Alignment Between Learning Objectives and Assessments in a Quantitative Literacy Course

Abstract

In this analysis, we examine how course assessment items were aligned with learning objectives in a quantitative literacy course at Michigan State University. The alignment analysis consisted of mapping assessment items to a list of operationalized learning objectives from the course. Our analysis shows how often the learning objectives are represented in assessment items, how often they are paired with other learning objectives, and how influential they are in contributing to a student's course grade. In addition, through comparisons across four assessment types (e.g., exams and homework), we show how each learning objective was assessed differently within each assessment type. The most frequently represented learning objectives in the particular course we studied concern the creation and interpretation of graphical representations; these learning objectives were assessed relatively evenly across the assessment types. However, those learning objectives often co-occurred with other objectives in assessment items, and the point values per item associated with these objectives were less than those for other objectives. Our study shows how quantitative literacy learning objectives can vary with assessment type in a course, and also provides numeracy scholars with an analysis technique suitable for use at their respective institutions.

Cover Page Footnote

Younggon Bae is a doctoral candidate in mathematics education at Michigan State University. His research focuses on curriculum design and technology for undergraduate mathematics courses.

Samuel Luke Tunstall is Director of the new Quantitative Reasoning and Skills Center at Trinity University in San Antonio, Texas. In addition to serving on the board of the National Numeracy Network, he is also Chair-Elect of the Mathematical Association of America's Special Interest Group in Quantitative Literacy. His research interests include quantitative literacy practices and the assessment of quantitative literacy.

Kathryn S. Knowles is a doctoral candidate in the mathematics education program at Michigan State University. Her research interests include quantitative literacy, K-12 mathematics curriculum, and instruction.

Rebecca L. Matz is an academic specialist in the Hub for Innovation in Learning and Technology at Michigan State University. Her research focuses on assessment and student success in introductory undergraduate STEM courses.

Introduction

With sustained emphasis on student success in general education mathematics courses over the past two decades, new courses designed to foster quantitative literacy and reasoning have proliferated in recent years (Gaze 2018). In designing a new course, it is necessary to address what learning goals students are expected to achieve and how the course will offer opportunities to achieve those goals (Mathematical Association of America 1994). In this regard, one way to examine the validity of a course design is to investigate how the course assessments align with the course learning goals, especially as intentional course alignment has been shown to positively impact students' approaches to learning (Wang et al. 2013). In particular, mapping course assessment items (e.g., mathematical tasks on exams or in-class group activities) to specific learning objectives can provide evidence about how the objectives are represented holistically across the course assessment materials. In coursework centered on quantitative literacy, attention to the alignment between learning goals and assessments is of particular importance, given that quantitative literacy is challenging to both operationalize and assess (Shavelson 2008; Boersma et al. 2011). Notwithstanding this challenge, as courses in quantitative literacy grow in scale—both in the form of large course sizes (Tunstall et al. 2016) and the distribution of uniform curricula (e.g., Quantway or Statway)[1] across institutions—there is a marked need for attention to the operationalization and assessment of learning objectives in such courses.

In this analysis, we examine how course assessment items are aligned with learning objectives in a quantitative literacy course at Michigan State University. Since its development (see Tunstall et al. 2016), this course has aimed to develop student competence in interacting with mathematical and statistical representations in real-life contexts and communicating their understanding through oral and written formats. The context-based and process-oriented aspects of quantitative literacy instantiated in this course align with extant literature of numeracy, quantitative literacy, and quantitative reasoning discussions (Steen 2001; Madison 2003; Karaali et al. 2016). With this perspective on a quantitative literacy course, the course design consisted of multiple contextual modules with different types of course assignments to foster students achieving the learning objectives.

This alignment analysis consisted of mapping assessment items to a list of operationalized learning objectives from the course syllabus and shows how often the learning objectives are represented in assessment items, how often they are paired with other learning objectives, and how influential they are in contributing

---

[1] See https://www.carnegiemathpathways.org/quantway/ for more information about offerings in relation to Carnegie Math Pathways.

to a student's course grade. In addition, through comparisons across four assessment types including exams, homework, projects, and in-class activities, we show how each learning objective was assessed differently within each assessment type. The most frequently represented learning objectives in the particular course we studied concern the creation and interpretation of graphical representations; these learning objectives were assessed relatively evenly across the assessment types. However, these learning objectives often co-occurred with other objectives in assessment items, and the point values per item associated with these objectives were less than those for other objectives. Some objectives, such as supporting an argument and critical analysis, are mostly assessed in particular types of assessments (projects and labs) due to the complexity of assessing students' writing and creation of graphical representations involved in those objectives. Given that those assessment types account for a smaller percentage in the overall course evaluation scheme than that for the exams, students' work toward those objectives did not impact their overall grades as much as they would under a different weighting scheme.

In sum, this study shows how quantitative literacy learning objectives can vary with assessment type in a course, and also provides numeracy scholars with an analysis technique suitable for use at their respective institutions.

Literature Review

Educators and educational researchers recognize that if components of an educational system (e.g., federal, state, and district policies; curriculum, assessments, and teaching practices) are not aligned, the system will be fragmented, confusing, and ineffective (University of Pennsylvania 1991; Newmann 1993; Spillane 1998). Alignment is not a new concept of study; researchers have used alignment studies to guide curriculum and instruction development (e.g., Hubball and Burt 2007; Coburn et al. 2016), validate assessment instruments (e.g., Webb 2002; Sundre and Thelk 2010; Wittstrom et al. 2010), and characterize the nature of assessments across disciplines (e.g., Momsen et al. 2013).

Indeed, since the course goals and assessment items (ideally) represent the instructor's expectations for student learning in their course designs, analysis of these two pieces is ostensibly informative. Researchers in various disciplines have evaluated how course assessment materials reflect course learning goals by mapping individual assessment items to a set of codes that represent the intended student learning of interest (Momsen et al. 2013; FitzPatrick et al. 2015; Laverty et al. 2016). For example, in the context of introductory biology and physics, Momsen et al. (2013) analyzed the level of cognitive demand represented in assessment items with respect to Bloom's Taxonomy, finding that most assessment items in their sample of high-stakes exams tested low-level cognitive skills. Laverty et al.

(2016) developed a protocol that characterizes the extent to which assessment tasks from college science courses can elicit evidence of student engagement with scientific practices, crosscutting concepts, and core disciplinary ideas (National Research Council 2012). A follow-up study evaluated change over time in the nature of assessments in introductory biology, chemistry, and physics courses with respect to these three dimensions (Matz et al. 2018).

In particular, some researchers have analyzed and judged the extent of content alignment between standards—a predetermined set of learning objectives—and assessment instruments in terms of how and to what extent the assessments cover the standards (Webb 2002; Wittstrom et al. 2010). For example, Sundre and Thelk (2010) conducted an alignment study to validate an assessment instrument for college students' quantitative literacy. They invited content experts from four different institutions to map each item from the instrument to their institutions' learning objectives for quantitative literacy. Given the high percentages (92–100%) of items that were successfully mapped, the researchers argued for the content validity of the instrument as well as its generalizability to those institutions.

Past literature demonstrates the usefulness of alignment studies in investigating course assessment materials, while providing various approaches and frameworks that reflect different purposes of the studies. In this study, we used a mapping analysis to examine the content alignment between course objectives and assessments in a quantitative literacy course.

Research Questions and Context

The purpose of this study is to examine alignment between course learning goals and assessment materials in a newly-developed quantitative literacy course (Math 101) at Michigan State University, a large, public, research-intensive university. More details about this course and its development are available in Tunstall et al. (2016). We describe the logistics of the course itself below, but note here that the Math 101 course is centered around context-specific modules that are subject to change from semester to semester. Since its development, the contexts have included: The World and Its People, Numbers in the Media, Science, and Health and Risk. In the modules, tools of quantitative reasoning are introduced as a means of better understanding issues or topics, rather than the reverse—an approach that has both benefits and limitations. Benefits include, among other things, that students tend to be engaged insofar as the contexts themselves are engaging, as well as that we cover various facts of life in which quantitative reasoning is often considered useful (Madison and Steen 2007). On the other hand, a limitation is that when contexts drive a course, it is difficult to define and then operationalize learning objectives that convey the mathematics and statistics being done. In some respects, this limitation constitutes the reason we arrived at this study.

Research Questions

With this context in mind, we conducted a course alignment analysis to address the following research questions: RQ1) What specific learning objectives were the most and least represented across the course assessments? and RQ2) How were the learning objectives represented differently within each type of assessment? We collected and analyzed the course syllabus and all the assessment materials used to determine the students' final grades. To examine how the assessment items reflected the learning objectives, all the items were coded with a set of Learning Objectives (LOs), operationalized from the original set of objectives in the course syllabus. This analysis shows how the LOs were represented in the assessments throughout the course in terms of exhaustiveness and any differential emphasis on particular objectives.

It is important to note that we—the authors of the paper—have each been involved in some way with the development and teaching of Math 101 since the inception of the course. In the context of this work, some of the authors contributed to the development of the learning objective base that has evolved as the course has also evolved; however, the instructor for the Fall 2017 semester (the term of interest in this study) designed the course syllabus and assessments. Our role here is not to critique or make normative claims about the design or enactment of the course, and we are not attempting to discern the instructor's intentions. Instead, our analysis centers on the relationship between learning objectives and their instantiation in the assessments. We also note that the course changes incrementally from semester to semester, so these analyses should not necessarily be considered indicative of the current course.

An additional important remark is that our approach to analyzing alignment between the course learning goals and assessments was based on the analysis of written documents, and that the scope of our analysis did not include teaching practices in the classroom or student performance on the assessments. In this study, we do not ask questions or make claims about what the students learned from this course and how the student learning took place; addressing these issues is not the purpose of the study. Rather we argue that this analysis, from a course design perspective, allows us to examine the extent to which the assessments in this relatively new course are aligned with the learning goals.

Context

In Fall 2017, 379 total students enrolled across 14 sections of Quantitative Literacy I (Math 101) with 22 to 30 students per section. The course offered two weekly class meetings, a large 80-minute lecture session, and a small 80-minute recitation session for each section. The instructor of the course provided a lecture for students on Tuesdays followed by a recitation session for each section on Thursdays in which the students met with their recitation leader to work on lab activities related

to the topics taught in the lecture of the week. In the recitation sessions, students collaborated in groups of three or four peers to solve problems and then turned in their worksheets for grading. A subset of students in the course was enrolled in enhanced sections, meaning they additionally attended a recitation session on Wednesdays between the lecture and the laboratory sessions; an analysis of the efficacy of these enhanced sections is reported in Matz and Tunstall (2018).

The larger course goals include approaching information with a numerical lens, emphasizing how mathematics informs real-life contexts, and communicating understanding through oral and written formats aligned with extant literature on numeracy, quantitative literacy, and quantitative reasoning (Steen 2001; Madison 2003; Karaali et al. 2016). Students enrolled in Math 101 are encouraged to approach numerical information presented to them in daily life with a critical lens. The contexts included within the course modules highlight how mathematical skills, such as mathematical modeling, may help with day-to-day demands of students' lives. In the module The World and Its People, for instance, students engage with Gapminder, a modeling tool, to consider important social questions about world health. Students also create their own mathematical models from real-world problems or health statistics to develop their mathematical skills and broaden their perspective.

In order for students to develop their understanding of mathematics in context, communication is a main feature of the course. Students are encouraged to verbally communicate, specifically during recitations, with classmates and instructors. Students work together in groups to read, interpret, and find solutions to real-world problems. At times, students present their understanding by using appropriate mathematical symbols, graphs, and language. In addition to verbal communication, written communication is also used within the course to assess student learning.

Written evidence of student learning was shown on four types of graded course assessments: (a) exams, (b) labs, (c) individual projects, and (d) homework assignments. Together, these four types of assessments accounted for 1,000 points in total which determined the course grade. Exams consisted of multiple choice, open-ended, and fill-in-the-blank questions. Labs specifically asked students to discuss topics from the previous lecture. The lab questions could be centered on social media posts, news articles, or other sources which related to a specific mathematical goal. For example, while discussing overpopulation and world health issues (e.g., access to safe drinking water and basic sanitation), students worked with ideas about proportions, absolute and relative measures, and graphical models. Individual projects provided students the opportunity to show what they know with additional flexibility. In one case, students chose a Millennium Development goal from a pre-existing list to create a story about various populations around the world and their progress on the selected goal. Many possible solutions for these individual projects exist, allowing students to interpret, think about, and present their

understanding in multiple ways. Homework assignments gave more guidance and practice about key concepts and skills covered within lectures. Each assignment consisted of review content along with a few "check-up" questions and links to real-life examples (e.g., TED talk video links, news articles, or statistical data).

Methods
Data Collection

We collected the Fall 2017 Math 101 course syllabus containing the course-level learning objectives, descriptions of the course assessments, and the grading scheme noting specific point allocations to each assessment type. Assessment materials that contributed to the final grade were also collected, including three written exams (two midterms and one cumulative final), ten worksheets for labs, instructions for two projects, and nine homework assignments.

Data Analysis

Assessment Items. Individual assessment items in the course materials were identified as the unit of analysis; that is, each question in the exams, labs, and homework was considered an individual item as written and was labeled as such in our materials. Some assessment items were naturally clustered (by the instructor) to address particular concepts or procedures—for example, a problem might contain three subparts labeled (a), (b), and (c). In these cases, we considered each subpart separately in applying the Learning Objective (LO) codes. Similarly, we identified the items in the project assessments based on the smallest unit of instruction (e.g., creating a presentation slide with a given set of information or writing a paragraph that responded to a question prompt). The numbers of identified individual items and corresponding point values from the four assessment types are detailed in Figure 1.
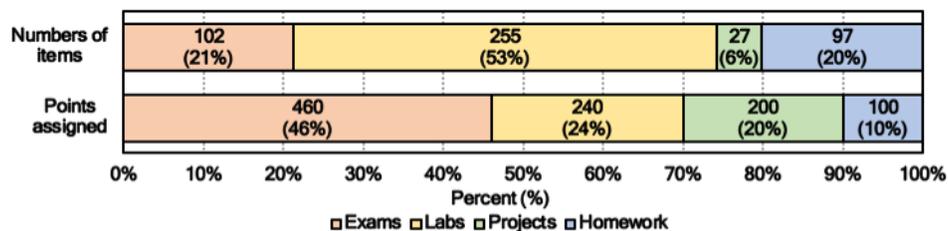

Figure 1. The numbers of assessment items and point distribution across the assessment types. In total, 481 items were identified across the four assessment types, representing the 1,000 total points for the course. The number of items and associated points varies with assessment type. Exams and projects contain fewer items relative to the associated points compared to labs and homework.

Learning Objectives. Iterative development of the LO codes began with operationalizing the original learning objective statements as they appeared in the

course syllabus, written by the course instructor. This operationalization by our research team consisted of both consolidating and dividing the original objectives so as to support coders in reliably interpreting and applying the LO codes. That is, in the original learning objectives we identified both a few cases where the same idea was stated in multiple learning objectives, and a few cases where one original learning objective seemed to contain multiple ideas (Fig. 2). We also added an "Other" LO code because we encountered a small number of assessment items that could not reasonably be coded with any of the existing objectives. Those items include converting between units of measurement and collecting background information on students' interests.
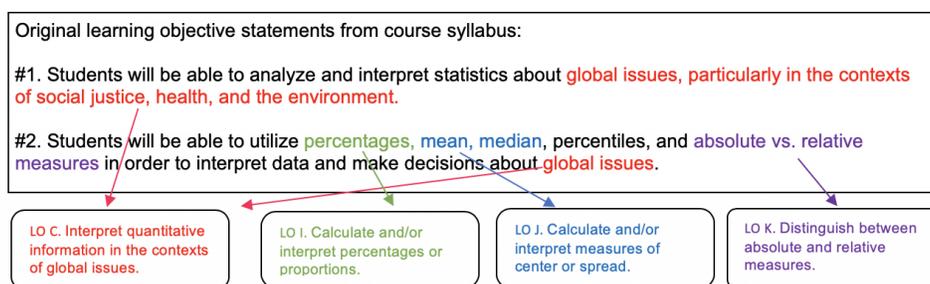
Original learning objective statements from course syllabus:

#1. Students will be able to analyze and interpret statistics about global issues, particularly in the contexts of social justice, health, and the environment.

#2. Students will be able to utilize percentages, mean, median, percentiles, and absolute vs. relative measures in order to interpret data and make decisions about global issues.

| LO C. Interpret quantitative information in the contexts of global issues. | LO I. Calculate and/or interpret percentages or proportions. | LO J. Calculate and/or interpret measures of center or spread. | LO K. Distinguish between absolute and relative measures. |

Figure 2. Examples of multiple LO codes in an original learning objective statement and an LO code derived from multiple statements. Learning objective #2 from the course syllabus contains multiple ideas, operationalized as LOs C (global issues), I (percentages and proportions), J (measures of center or spread), and K (absolute and relative measures). LO C is also contained in learning objective #1 on the syllabus.

As consistent coding requires categories that can be reliably distinguished (Saldaña 2009), this process was necessary, though we note that, in sum, only minor modifications had to be made to the original list of learning objectives; many objectives were already distinctive in their original form. Having identified the mathematical activities and reasoning included in the statements, we generated the final list of LOs (without a specific order implying importance) for coding with brief descriptions (Appendix 1) as well as a list of example assessment items by LO code (Appendix 2). The codes were finalized after three iterations of pilot coding on random samples conducted by three coders (Hruschka et al. 2004). During all coding rounds, raters had access to the group's notes about all items that had been previously coded.

Coding and Inter-Rater Reliability. For each item, the coder applied the most accurate code(s) describing the primary mathematical activity or reasoning required in the task. Two codes at most could be applied to a single item if they were both determined to be salient in describing the nature of the task. In one example problem from a lab activity (Fig. 3), for instance, students are asked to make and

support an argument using quantitative information (LO A) requiring a valid interpretation of a somewhat unusual infographic (LO H). This process of simultaneous coding (Saldaña 2009) was used because we identified numerous similar items that appeared to rely on more than one objective (Sikorskii et al. 2011), but the limit of two codes was also a practical choice to help maintain reliability between different coders.
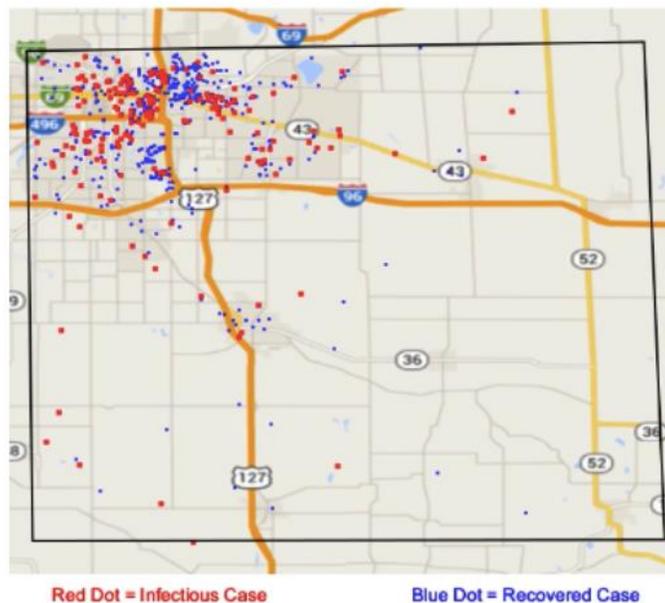
Figure 3. An example assessment item from a lab activity that was coded with two LOs: A (supporting an argument) and H (infographics). Students are asked to interpret the interactive simulation of a measles outbreak on a map as well as make an argument with supporting quantitative information.

To examine our inter-rater reliability in coding, the code applications from two independent coders were compared. For each item, an agreement on the code application was identified if one or more concurrent LO code(s) was applied by both coders. After three rounds of preliminary coding, discussion, and refining the codebook, the percentage agreement on a random sample of 20 items (representing

all the different types of assessment) between the two coders was 100%, and all the assessment items were subsequently coded with the LO codebook and protocol developed by the researchers. While establishing inter-rater reliability with percent agreement does not correct for agreement by chance and thus always overestimates the actual agreement between raters (Hallgren 2012), we argue that the number of possible codes (19) is large enough to make agreement by chance very small relative to our observed agreement. After inter-rater reliability was established, one coder was responsible for coding the remainder of the dataset.

Measures. Based on the LO code applications to the assessment items, three separate measures, described below, were calculated to determine how each LO was represented: (a) the number of items representing each LO code, (b) given the allowance for double-coding items, an adjusted number of items representing each LO code, and (c) points assigned to items representing each LO code.

First, the number of items representing each LO code was determined by the number of all the items that reflected the code. Since there were 217 items with two code applications, the sum of the measures from all the codes is not the total number of items (N=481) but rather equal the total number of code applications (N=698). This measure indicates on how many items in the course the students were asked to use their knowledge of the given LO.

Second, the adjusted number of items representing each LO code was determined by using different weights for items with two code applications. To consider the double-counting of the items with two code applications, each of the two codes from an item was counted as the half for evaluating the adjusted numbers of items. The sum of this measure from all LO codes is, therefore, equal to the total number of items (N=481). Comparing the number of items and the adjusted number of items representing an LO code shows the extent to which each LO was combined with another LO in the course assessments.

Last, the points assigned to items representing each LO code indicates the sum of all the points from the items that represent the code. For an item with two code applications, each code was assigned half of the points. This measure indicates the extent to which student performance on each LO contributed to their final grade.

Limitations

We do not claim that the learning objectives are all written at the same level. For example, some objectives rely on relatively specific skills (e.g., calculating proportions and creating graphs) compared to others that address broader concepts and student reasoning (e.g., supporting an argument with quantitative information). Additionally, for the items with two code applications, we divided the points in half for each code assuming that the two objectives were equally represented in the item, though this distribution is unlikely to be exactly equal for many items. Further, we

did not include the small number of optional learning activities that students had access to as part of the course, such as participating in weekly online discussion forums or a pre/post cognitive assessment for learning gains. Finally, we note that the labs were often purposefully designed to take longer than a single class meeting to complete, but our point-weighting scheme assumes that students would have completed the entirety of each lab.

Analysis

This alignment analysis shows how the learning objectives are represented across the different assessment types in terms of three measures, (a) the number of assessment items representing each objective, (b) the adjusted number of assessment items representing each objective, and (c) the points assigned to items representing each objective. In the following sections, we present the representation of LOs by assessment type according to these three measures and evaluate which LOs are most and least represented. We then discuss how the representation of LOs varied across assessment types by comparing the measures to the percentages of the course evaluation scheme for each assessment type under the assumption that the representation of LOs is independent of assessment types.

RQ1: Representation of LOs Across All Assessments

To examine how each LO code is represented in the course assessments, we evaluated the three measures for each code and compared them to identify the most and least represented LOs in the assessments (Table 1). In addition, comparing the three measures shows how often the codes co-occurred with other codes in a single assessment item and the extent to which the LOs had influence in determining overall student grades according to how often they appeared in the assessment items and their relative point values. By all three measures, LOs F (graphs) and H (infographics) are the most represented learning objectives across the course assessments; that is, these learning objectives are assessed most frequently in the course and account for the most points in the assessments. LOs M (expected value) and P (awareness of limitations) are the least represented objectives by all three measures.

Table 1
The Numbers of Assessment Items, the Adjusted Numbers of Assessment Items, and the Points Assigned for Each LO Code

LO (Shortened labels)	Number of Items (%)	Adjusted number of items (%)	Points assigned (%)
A. Supporting an argument	20 (4)	14 (3)	54 (5)
B. Quantitative assumptions	10 (2)	7.5 (2)	20 (2)
C. Global issues	49 (10)	27 (6)	52 (5)
D. Critical analysis	29 (6)	21.5 (4)	77 (8)
E. Formula	13 (3)	9.5 (2)	13 (1)
F. Graphs	105 (22)	72 (15)	110 (11)
G. Tables	22 (5)	14.5 (3)	21 (2)
H. Infographics	101 (21)	61.5 (13)	119 (12)
I. Percentages and proportions	72 (15)	51.5 (11)	95 (10)
J. Measures of center or spread	63 (13)	48.5 (10)	113 (11)
K. Absolute and relative measures	28 (6)	17 (4)	22 (2)
L. Stochastic phenomenon	36 (7)	29 (6)	73 (7)
M. Expected value	3 (1)	3 (1)	3 (0)
N. Bayes theorem	53 (11)	40 (8)	72 (7)
O. Prediction	39 (8)	29 (6)	73 (7)
P. Awareness of limitations	4 (1)	2.5 (1)	4 (0)
Q. Correlation and causation	29 (6)	15 (3)	32 (3)
R. Interpolation and extrapolation	15 (3)	11 (2)	25 (3)
S. Other	7 (1)	7 (1)	23 (2)
Total	698 (145)	481 (100)	1000 (100)

Frequency of LO Representation in Assessment Items. The number of items representing each LO code shows how often students are asked to use their knowledge or skills of the LO. The number of items for each LO code ranges from 3 (1%) to 105 (22%), and the median is 29 (6%). Since the items with two codes applied were counted twice in this measure, the sum of percentages from all the codes is 145%, reflecting that 217 of the 481 items were coded with two objectives. This measure for each LO code shows what percent of items reflect the LO regardless of whether the item was coded with one or two LOs. Comparison across different codes shows what particular LOs are assessed relatively more or less in the assessment items. For example, the most represented LO codes by this measure are LOs F (graphs) and H (infographics), which were applied in 105 and 101 items, respectively, among the total 481 items in the course, meaning each of the LOs was addressed in more than 20% of the course assessment items. On the other hand,

LOs M (expected value) and P (awareness of limitations) appeared in 3 and 4 items, respectively, each constituting less than 1% of the assessment items in the course.

Pairwise Appearance of LOs. The adjusted number of items representing each LO code indicates how often the code co-occurs with other codes in a single assessment item. Because the items with two code applications were counted double in the first measure, the numbers and percentages decrease overall in this adjusted measure. The adjusted numbers for each LO code range from 2.5 (1%) to 72 (15%), and the median is 17 (4%). The most and the least represented LO codes by this adjusted measure were the same as the first measure. However, the adjusted numbers of items for LOs F (graphs) and H (infographics), which are the most represented codes according to both measures, show significant decreases by the adjusted measure, implying that these LOs most often co-occurred with other LOs. For example, out of 105 items coded with the LO F, 66 items included another code while only 39 items were coded exclusively with LO F.

Influence of LOs on Final Course Grades. The points assigned to the LO codes show the point values assigned to items representing each LO code, which in turn implies how student performance on each LO influenced their course grade. The points for each LO code (out of 1,000 total course points) ranges from 3 (0%) to 119 (12%), and the median is 52 (5%). The most represented LOs by this measure are again F (graphs) and H (infographics), the same as for the first two measures, but also includes LO J (measures of center or spread), which shows a relatively small drop in percentage from the number of items to the points assigned. The least represented codes are again LOs M (expected value) and P (awareness of limitations) which each account for no more than 1% of the entire points across the course.

The comparison between the points assigned and the adjusted numbers of items for each LO indicates how the items representing each LO are valued relative to how often the objective was assessed in the assessment materials. The percent of the points assigned for nine LOs (A, B, D, J, L, O, Q, R, and S) increased from the adjusted number of items whereas the remaining ten LOs (C, E, F, G, H, I, K, M, N, and P) showed decreases. For instance, for LOs A (supporting an argument) and D (critical analysis), the percent of course points was greater than the percent of number of items representing the code, indicating that these two LOs were highly valued. However, the percent of points assigned to LO F (graphs, 11%) decreased from percent of the number of items coded with LO F (15%). This implies that student work on items coded with LO F is relatively less valued than that for LOs A and D, though LO F is the most frequent objective that students were asked about in their assessment materials (Fig. 4).
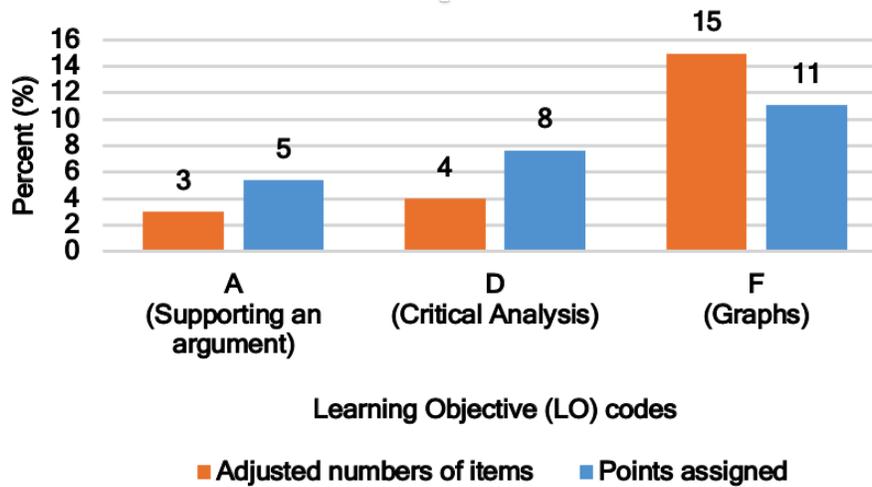
Figure 4. Differences between the adjusted number of assessment items and the points assigned for LO codes A (supporting an argument), D (critical analysis), and F (graphs).

RQ2: Differences in Representation of LOs by Assessment Type

To examine differentials in the distribution of LOs by assessment type, we evaluated the percent of points assigned to each LO within each assessment type compared to the total points assigned to the LO (Fig. 5). Under the assumption that the LOs are not preferred by a particular assessment type, we expected that the distribution of the points assigned to each LO would be exactly the same as the entire course assessment distribution. However, the analysis reveals multiple patterns of distribution as shown in Figure 5 that are different from the overall course assessment distribution. Some objectives were represented mostly in exams while others were assessed in the more open-ended project and lab assessments.

Figure 6 shows the overall course assessment distribution (dotted line) and the distribution of points for three LOs that are representative of three patterns identified in this analysis. First, LO F (graphs) represents a pattern that is close to the overall course assessment distribution. The distribution of points assigned to LOs F, G (tables), and H (infographics) shows a similar pattern to the overall course assessment distribution. That is, 46% of the overall course points come from exams, and roughly 46% of the points associated with LO F also come from exams. For these LOs, more points were assigned to each assessment type in the order of exams, labs, projects, and homework, which is same as the course distribution.
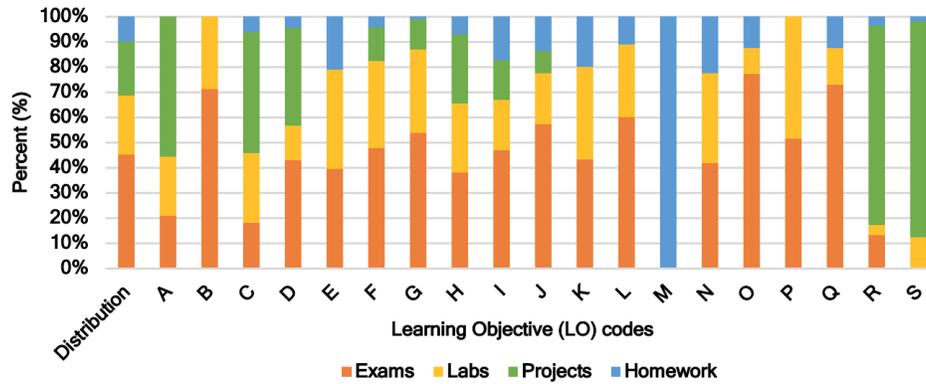
Figure 5. The distribution of points for each LO code across the assessment types, showing how the points assigned to each LO are distributed across the different assessment types.
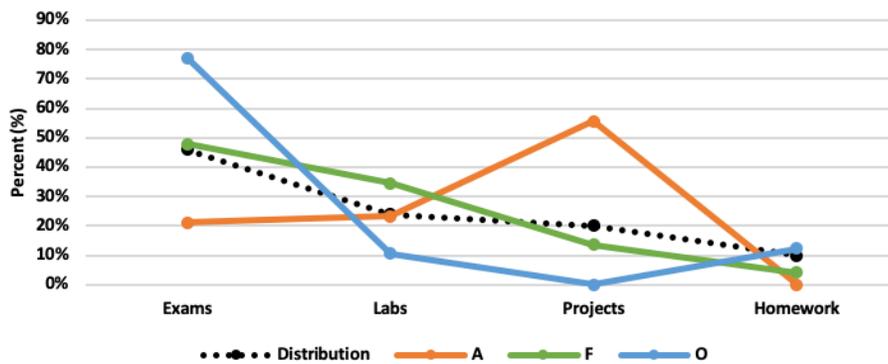

Figure 6. The distribution of points assigned for LOs A (supporting an argument), F (graphs), and O (prediction) across the assessment types that exemplify three different general patterns of distribution.

Other LOs including A (supporting an argument), C (global issues), R (interpolation and extrapolation), and S (other) showed a different pattern of distribution in which points were assigned more in projects and less in exams compared to the overall course distribution (see LO A in Fig. 6). For example, the comparison indicates that more than 50% of the points assigned to LO A were found in projects whereas about 20% of the points were assigned in exams, implying that the assessment of LO A in this course preferentially occurred in projects compared to other assessment types. This difference in distribution shows that projects were the most significant assessment type in this course for reflecting student achievement in these LOs.

Finally, more than 70% of points assigned to LOs B (quantitative assumptions), O (prediction), and Q (correlation and causation) were identified in exams whereas relatively fewer points were found in labs and projects compared to the overall course distribution (see LO O in Fig. 6). This distribution indicates that the influence of exams on the evaluation of student achievement for these LOs was more significant than would be expected under the assumption that the assessment type and learning objective are unrelated.

Discussion and Implication

The analysis of the three measures for representation of LOs shows that the most represented LOs among all three measures were to create and interpret mathematical representations of quantitative information, including graphs (LO F) and infographics (LO H). Further, comparisons of the three measures and the assessment types imply more than just the dominance of their representations in the assessments. LOs F and H were paired with other objectives in the assessment items more than any other and found across all assessment types. However, individual items representing those objectives were less valued in determining student grades than other objectives.

Core Objectives in Assessments: Creating and Interpreting Graphical Representations

The mathematical tasks of creating and interpreting both graphs and infographics were frequently integrated with other objectives at the individual item level as well as included across all assessment types. These LOs show substantial drops in the adjusted numbers of items from the numbers of the items, indicating that they frequently co-occurred with other objectives in the individual items. For instance, 63% (66 of 105) of the items representing LO F (graphs) were also coded with another objective, as were 78% (79 of 101) of items representing LO H (infographics). These results demonstrate the crucial role of mathematical and statistical representations in context-based course design for quantitative literacy and reasoning courses as discussed in the extant literature (e.g., Madison 2014). Second, the comparison by assessment types shows that these two LOs were assessed in each assessment type proportionate to the overall course evaluation distribution. This finding indicates that the representation of these LOs was not biased by assessment type, and that the students engaged with the activities of creating and interpreting graphs and infographics consistently across the exams, labs, projects, and homework.

In that sense, creating and interpreting graphical representations were the core objectives in the course assessments of Math 101. Given the prevalence of interaction with multiple representations across the course assessments, these

findings imply that it is important for students' success in this course to develop those abilities and apply them regularly in complex practices such as collecting quantitative information from real-life contexts and supporting arguments with visual representations.

Assessing Learning Objectives for Effective Communication

In addition, the analysis shows that the learning objectives were not all equally distributed across different types of assessments in this course. Though we do not claim a stable correlation between particular assessment types and learning objectives that would generalize to other courses, our analysis implies that designing course assessments aligned with multiple learning objectives requires close examination for possible (unintentional) over- and underrepresentation of objectives within different assessment types. In this course assessment distribution for Math 101, more points are assigned to exams (46%) than labs (24%), projects (20%), and homework (10%). Some LOs that are represented mostly in exams but rarely found in homework (e.g., LOs B, O, and Q) had more influence on determining student final grades than other LOs that are represented less often in exams but mostly in other types of assessments (e.g., LOs A, C, R, and S). Such LOs in this case were perhaps more conducive to being graded quickly in the form of a closed-response exam or automated-grading homework system; LO A (for example), "Analyze and use quantitative information to support an argument," is likely to be assessed through writing (which in this course would be in labs or projects).

These findings imply that course assessment design should include attention to differences in the representation of learning objectives by assessment type as well as the distribution of points assigned for each type, as they can reveal an over- or underrepresentation of particular objectives in student evaluation. Indeed, if a higher weight is allotted to a specific assessment type (e.g., closed form assessments), then inevitably learning objectives conducive to that assessment type will be those most "represented" in the course (e.g., those involving calculation). In light of existing literature concerning the importance of argumentation and writing in the assessment of quantitative literacy and reasoning (e.g., Grawe 2011), this finding accentuates the importance of awareness of what is gained and lost as we use different assessment types, especially as courses grow in scale.

Implications for Instructors and Developers of QL Courses

The findings of this study provide actionable implications, especially for instructors and course developers in other institutions where intended learning goals and pedagogical approaches are similar to Math 101. First, it is important to identify core learning objectives among others and, more importantly, to understand how

those core objectives are instantiated in the existing course assessments. In support of student success, courses would benefit from an explicit emphasis on core objectives that are required in many different contexts and assignments throughout the course. Though multiple contextual modules in such courses are not driven by hierarchical structure of mathematical content, instructors may attend to core objectives and emphasize applications to different contexts and practices. Second, determining the course evaluation distribution across the multiple course assessment types requires close examination of the extent to which each assessment type represents a variety of quantitative literacy learning objectives. Given the different nature of objectives as shown in this study, it may not be possible to distribute all the objectives evenly across all the assessment types. Rather, we recommend finding an appropriate distribution that sufficiently evaluates student performance when considering the holistic assessment picture of the entire course (and perhaps sequence of courses). In many institutions, quantitative literacy courses have been developed to provide an alternative path from traditional developmental courses and to better serve students who suffered from unpleasant experiences in procedure-oriented and exam-driven mathematics courses. If the overall course evaluation prefers student performance on some objectives that traditionally associated with closed-ended exam problems, even if those objectives are core elements, the student loses the opportunity to be assessed by other objectives.

Conclusion

This analysis of alignment between learning objectives and course assessments provides information about how often students were asked to engage with the course-level objectives, the overlap between objectives in assessment items, the extent to which each objective influenced students' final course grade, and the distribution of objectives by assessment type. The most highly represented learning objectives were creating and interpreting graphs (F) and infographics (H), and the analyses indicate that these objectives often co-occurred with other objectives in assessment items. While these objectives were assessed relatively evenly across the four assessment types, the point values associated with these objectives were less than would be expected if point values and frequencies associated with the objectives were proportionate.

In contrast, some learning objectives related to effective communication—supporting an argument using quantitative information and critically analyzing misleading information—were valued relatively more in determining final course grades compared to how often they were presented in assessments. Given the multiple objectives represented differently by assessment type, this study (as discussed above) calls attention to the influence of assessment types on potential

over- and underrepresentation of learning objectives in evaluating student grades. In sum, we suggest that this method of alignment analysis is useful for providing feedback on the design of course assessments, unpacking the nature of learning objectives represented in the assessments, and generally examining the validity of a course design.

Acknowledgements

References

Boersma, Stuart, Caren Diefenderfer, Shannon W. Dingman, and Bernard L. Madison. 2011. "Quantitative Reasoning in the Contemporary World, 3: Assessing Student Learning." *Numeracy* 4(2): Article 8. https://doi.org/10.5038/1936-4660.4.2.8

Coburn, Cynthia E., Heather C. Hill, and James P. Spillane. 2016. "Alignment and Accountability in Policy Design and Implementation: The Common Core State Standards and Implementation Research." *Educational Researcher* 45(4): 243–251. https://doi.org/10.3102/0013189X16651080

FitzPatrick, Beverly, John Hawboldt, Daniel Doyle, and Terri Genge. 2015. "Alignment of Learning Objectives and Assessments in Therapeutics Courses to Foster Higher-Order Thinking." *American Journal of Pharmaceutical Education* 79(1): Article 10. https://doi.org/10.5688/ajpe79110

Gaze, Eric. 2018. "Quantitative Reasoning: A Guided Pathway from Two-to-Four-Year Colleges." *Numeracy* 11(1): Article 1. https://doi.org/10.5038/1936-4660.11.1.1

Grawe, Nathan D. 2011. "Beyond Math Skills: Measuring Quantitative Reasoning in Context." *New Directions for Institutional Research* 149: 41–52. https://doi.org/10.1002/ir.379

Hallgren, Kevin A. 2012. "Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology* 8(1): 23. https://doi.org/10.20982/tqmp.08.1.p023

Hruschka, Daniel, J., Deborah Schwartz, Daphne Cobb St. John, Erin Picone-Decaro, Richard A. Jenkins, and James W. Carey. 2004. "Reliability in Coding Open-Ended Data: Lessons Learned from HIV Behavioral Research." *Field Methods* 16(3): 307–331. https://doi.org/10.1177/1525822X04266540

Hubball, Harry, and Helen Burt. 2007. "Learning Outcomes and Program-Level Evaluation in a Four-Year Undergraduate Pharmacy Curriculum." *American Journal of Pharmaceutical Education* 71(5): Article 90. https://doi.org/10.5688/aj710590

Karaali, Gizem, Edwin H. Villafane Hernandez, and Jeremy A. Taylor. 2016. "What's in a Name? A Critical Review of Definitions of Quantitative Literacy, Numeracy, and Quantitative Reasoning." *Numeracy* 9(1): Article 2. https://doi.org/10.5038/1936-4660.9.1.2

Laverty, James T., Sonia M. Underwood, Rebecca L. Matz, Lynmarie A. Posey, Justin H. Carmel, Marcos D. Caballero, Cori L. Fata-Hartley, Diane Ebert-May, Sarah E. Jardeleza, and Melanie M. Cooper. 2016. "Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol." *PloS one* 11(9): e0162333. https://doi.org/10.1371/journal.pone.0162333

Madison, Bernard L. 2003. "The Many Faces of Quantitative Literacy." In *Quantitative Literacy: Why Numeracy Matters for Schools and Colleges*, edited by Bernard L. Madison and Lynn A. Steen, 3–6. National Council on Education and the Disciplines.

Madison, Bernard L. 2014. "How Does One Design or Evaluate a Course in Quantitative Reasoning?" *Numeracy* 7(2): Article 3. https://doi.org/10.5038/1936-4660.7.2.3

Madison, Bernard L., and Lynn Arthur Steen, eds. 2008. *Calculation Vs. Context: Quantitative Literacy and Its Implications for Teacher Education: June 22–24, 2007, Wingspread Conference Center, Racine, Wisconsin*. Washington, DC: Mathematical Association of America.

Matz, Rebecca L., Cori L. Fata-Hartley, Lynmarie A. Posey, James T. Laverty, Sonia M. Underwood, Justin H. Carmel, Deborah G. Herrington, Ryan L. Stowe, Marcos D. Caballero, Diane Ebert-May, and Melanie M. Cooper. 2018. "Evaluating the Extent of a Large-Scale Transformation in Gateway Science Courses." *Science Advances* 4(10), eaau0554. https://doi.org/10.1126/sciadv.aau0554

Matz, Rebecca L., and Samuel L. Tunstall. 2018. "Investigating Embedded Remediation as a Pathway for Equitable Access to Quantitative Literacy and College Algebra." Manuscript submitted for publication.

Mathematical Association of America. 1994. *Quantitative Reasoning for College Graduates: A Complement to the Standards*. http://www.maa.org/programs/faculty-and-departments/curriculum-department-guidelines-recommendations/quantitative-literacy/quantitative-reasoning-college-graduates

Momsen, Jennifer, Erika Offerdahl, Mila Kryjevskaia, Lisa Montplaisir, Elizabeth Anderson, and Nate Grosz. 2013. "Using Assessments to Investigate and Compare the Nature of Learning in Undergraduate Science Courses." *CBE— Life Sciences Education* 12(2): 239–249. https://doi.org/10.1187/cbe.12-08-0130

National Research Council. 2012. *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13165

Newmann, Fred M. 1993. "Beyond Common Sense in Educational Restructuring: The Issues of Content and Linkage." *Educational Researcher* 22(2): 4–22. https://doi.org/10.3102/0013189X022002004

Saldaňa, Johnny. 2009. *The Coding Manual for Qualitative Researchers*. London: SAGE Publications Ltd.

Shavelson, Richard J. 2008. "Reflections on Quantitative Reasoning: An Assessment Perspective." In *Calculation vs. Context: Quantitative Literacy and Its Implications for Teacher Education*, edited by Bernard L. Madison and Lynn A. Steen, 27–47. Mathematical Association of America.

Sikorskii, Alla, Vince Melfi, Dennis Gilliland, Jennifer Kaplan, and Suzie Ahn. 2011. "Quantitative Literacy at Michigan State University, 1: Development and Initial Evaluation of the Assessment." *Numeracy* 4(2): Article 5. https://doi.org/10.5038/1936-4660.4.2.5

Spillane, James P. 1998. "State Policy and the Non-Monolithic Nature of the Local School District: Organizational and Professional Considerations." *American Educational Research Journal* 35(1): 33–63. https://doi.org/10.3102/00028312035001033

Steen, Lynn Arthur. 2001. *Mathematics and Democracy: The Case for Quantitative Literacy*. Princeton, NJ: National Council on Education and the Disciplines.

Sundre, Donna L., and Amy D. Thelk. 2010. "Advancing Assessment of Quantitative and Scientific Reasoning." *Numeracy* 3(2): Article 2. https://doi.org/10.5038/1936-4660.3.2.2

Tunstall, Samuel L., Vincent Melfi, Jeffrey C. Craig, Richard Edwards, Andrew Krause, Bronlyn Wassink, and Victor Piercey. 2016. "Quantitative Literacy at Michigan State University, 3: Designing General Education Mathematics Courses." *Numeracy* 9 (2): Article 6. https://doi.org/10.5038/1936-4660.9.2.6

University of Pennsylvania. 1991. Putting the Pieces Together: Systems for School Reform. CPRE Policy Briefs. https://repository.upenn.edu/cpre_policybriefs/63/

Wang, Xiaoyan, Yelin Su, Stephen Cheung, Eva Wong, and Theresa Kwong. 2013. "An Exploration of Biggs' Constructive Alignment in Course Design and Its Impact on Students' Learning Approaches." *Assessment & Evaluation in Higher Education* 38(4): 477-491. https://doi.org/10.1080/02602938.2012.658018

Webb, Norman L. 2002. *Alignment Study in Language arts, Mathematics, Science, and Social Studies of State Standards and Assessments for Four States*. Washington, DC: Council of Chief State School Officers.

Wittstrom, Kristina, Catherine Cone, Krista Salazar, Rucha Bond, and Karen Dominguez. 2010. "Alignment of Pharmacotherapy Course Assessments with Course Objectives." *American Journal of Pharmaceutical Education* 74(5): Article 76. https://doi.org/10.5688/aj740576

Appendices

Appendix 1
Learning Objectives
A. Supporting an argument: Analyze and use quantitative information to support an argument.
B. Quantitative assumptions: Recognize and evaluate quantitative assumptions in real-world sources.
C. Global issues: Interpret quantitative information in the contexts of global issues such as social justice, health, and the environment.
D. Critical analysis: Recognize when statistics and graphs in the media are presenting results in a misleading manner.
E. Mathematical representations: Create or interpret formulas as mathematical representations of information.
F. Mathematical representations: Create or interpret graphs (scatter plots, line and bar graphs) as mathematical representations of information.
G. Mathematical representations: Create or interpret tables as mathematical representations of information.
H. Mathematical representations: Create or interpret infographics (different from typical graphs) as mathematical representations of information.
I. Descriptive statistics: Calculate or interpret percentages or proportions.
J. Descriptive statistics: Calculate or interpret measures of center or spread such as mean, median, and standard deviation.
K. Descriptive statistics: Distinguish between absolute and relative measures of information, recognizing their benefits and limitations.
L. Reasoning with probability: Perform and interpret a simulation to explore a stochastic phenomenon.
M. Reasoning with probability: Use probability to calculate and interpret the expected value of a discrete random variable.
N. Reasoning with probability: Use Bayes Theorem to calculate and understand risk associated with a particular phenomenon.
O. Prediction: Make predictions about quantitative situations using mathematical models.
P. Awareness of limitations: Recognize that mathematical and statistical methods have limits.
Q. Correlation and causation: Recognize the difference between correlation and causation.
R. Interpolation and extrapolation: Interpret and predict the reasonableness of estimates obtained from interpolation and extrapolation.
S. Other

Appendix 2
Learning Objectives with Corresponding Exemplar Items

A. Supporting an argument: Analyze and use quantitative information to support an argument.	How is it possible that there is a global average of two children per woman, yet the population is still growing?
B. Quantitative assumptions: Recognize and evaluate quantitative assumptions in real-world sources.	If the population in a town has been increasing by 3% each year since 2010, when the population was just 15,000 individuals, write a formula for the population of the town since 2010. Discuss what would need to be true for the formula to yield an accurate prediction in the year 2030.
C. Global issues: Interpret quantitative information in the contexts of global issues such as social justice, health, and the environment.	How is it possible that there is a global average of two children per woman, yet the population is still growing?
D. Critical analysis: Recognize when statistics and graphs in the media are presenting results in a misleading manner.	Suppose a friend tells you that she is on a new diet, and—according to its creators—individuals have lost an average of 20% of their body weight. If this information is true, does this mean that the diet would be effective for you? Provide several mathematical reasons for why or why not.
E. Mathematical representations: Create or interpret formulas as mathematical representations of information.	If the population in a town has been increasing by 3% each year since 2010, when the population was just 15,000 individuals, write a formula for the population of the town since 2010. Discuss what would need to be true for the formula to yield an accurate prediction in the year 2030.

F. Mathematical representations: Create or interpret graphs (scatter plots, line and bar graphs) as mathematical representations of information.

The table below shows the number of accidents, fatalities, and hours flown for general aviation. Use the table to create a bar graph that makes an argument of your choice about whether aviation has become safer since 2003.

Years	Accidents	Fatalities	Hours flown (millions)
2003	1501	707	24.1
2008	1553	476	22.6

G. Mathematical representations: Create or interpret tables as mathematical representations of information.

The table below shows the number of accidents, fatalities, and hours flown for general aviation. Use the table to create a bar graph that makes an argument of your choice about whether aviation has become safer since 2003.

Years	Accidents	Fatalities	Hours flown (millions)
2003	1501	707	24.1
2008	1553	476	22.6

H. Mathematical representations: Create or interpret infographics (different from typical graphs) as mathematical representations of information.	Using your group's work from the lab on contraceptive measures, create an infographic that makes the case for the use of a specific form of contraceptive (e.g., birth control) that you believe is most effective. Include at least two distinct graphs, three statistics, and four pieces of general information to make your case.
I. Descriptive statistics: Calculate or interpret percentages or proportions.	Over the past three decades, as Brazil's economic growth has accelerated, the nation has tackled issues surrounding income inequality. Those in the highest income bracket of the country take in about 13% of the country's total income; the bottom 50% of the nation takes in roughly the same amount. Given that its population in 2015 was roughly 200 million people, how many people are in the top 1% of all of Brazil's income earners?
J. Descriptive statistics: Calculate or interpret measures of center or spread such as mean, median, and standard deviation.	Consider the top 1% of income earners in the U.S. Would it be reasonable to report the mean income of those individuals to describe the typical person in the 1%? Explain why or why not.
K. Descriptive statistics: Distinguish between absolute and relative measures of information, recognizing their benefits and limitations.	In conversation, a friend explains the U.S. plans to give Tanzania 587.7 million dollars in 2015 with the implication that this is too much of the federal budget. Without doing any further research do you agree that this amount of aid is "too much," why or why not?
L. Reasoning with probability: Perform and interpret a simulation to explore a stochastic phenomenon.	Suppose that with a certain disease, there is a one in six chance that a person will not live beyond five years after diagnosis of the disease. Use dice to determine how many people—among a group of 30 individuals—would survive after five years since their individual diagnoses. Does the number of people from your rolling of dice match the expected number? Why or why not?
M. Reasoning with probability: Use probability to calculate and interpret the expected value of a discrete random variable.	Suppose you take out a fire insurance policy on your home. The annual premium is $300. In case of fire, the insurance company will pay you $200,000, with no deductible. The probability of a house fire in your area is 0.0002. What is the expected value of the amount of money you pay/receive from the plan?
N. Reasoning with probability: Use Bayes Theorem to calculate and understand risk associated with a particular phenomenon.	You go to see the doctor about a growth. The doctor selects you at random to have a blood test for a certain disease, which is currently suspected to affect 1 in 10,000,000 people in the U.S. The test is 99% accurate, in the sense that the probability of a false positive is 1%. The probability of a false negative is zero. If you test positive, what is the probability that you have that disease?
O. Prediction: Make predictions about quantitative situations using mathematical models.	For a given brand of condom, users are told that with proper use, the condom is effective 99.8% of the time. Among a group of 1,000 undergraduates who use the condom once, how many would we expect to have an unplanned pregnancy?

P. Awareness of limitations: Recognize that mathematical and statistical methods have limits.	Suppose that with a certain disease, there is a one in six chance that a person will not live beyond five years after diagnosis of the disease. Use dice to determine how many people—among a group of 30 individuals—would survive after five years since their individual diagnoses. Does the number of people from your rolling of dice match the expected number? Why or why not?
Q. Correlation and causation: Recognize the difference between correlation and causation.	Find two indicators on the site, Gapminder.org, that are positively correlated with one another. Explain potential reasons for this relationship.
R. Interpolation and extrapolation: Interpret and predict the reasonableness of estimates obtained from interpolation and extrapolation	In the previous problem, you projected a future population based on current and past trends. What is this process called?
S. Other	What is the distinction between the sensitivity and specificity of a medical test?