
2015

The Levels of Conceptual Understanding in Statistics (LOCUS) Project: Results of the Pilot Study

Douglas Whitaker

University of Florida, whitaker@ufl.edu

Steven Foti

University of Florida, fotisj@ufl.edu

Tim Jacobbe

University of Florida, jacobbe@coe.ufl.edu

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



Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Science and Mathematics Education Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

Whitaker, Douglas, Steven Foti, and Tim Jacobbe. "The Levels of Conceptual Understanding in Statistics (LOCUS) Project: Results of the Pilot Study." *Numeracy* 8, Iss. 2 (2015): Article 3. DOI: <http://dx.doi.org/10.5038/1936-4660.8.2.3>

Authors retain copyright of their material under a [Creative Commons Non-Commercial Attribution 4.0 License](#).

The Levels of Conceptual Understanding in Statistics (LOCUS) Project: Results of the Pilot Study

Abstract

The Levels of Conceptual Understanding in Statistics (LOCUS) project (NSF DRL-111868) has created assessments that measure conceptual (rather than procedural) understanding of statistics as outlined in GAISE Framework (Franklin et al., 2007, Guidelines for Assessment and Instruction in Statistics Education, American Statistical Association). Here we provide a brief overview of the LOCUS project and present results from multiple-choice items on the pilot administration of the assessments with data collected from over 3400 students in grades 6-12 across six states. These results help illustrate students' understanding of statistical topics prior to the implementation of the Common Core State Standards. Using the four components of the statistical problem-solving process outlined in the GAISE Framework, middle and high school students tended to perform better on questions assessing Formulate Questions and Collect Data; both groups also tended to perform relatively worse on the Analyze Data questions.

Keywords

statistics education research, statistical literacy, assessment, conceptual understanding

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Cover Page Footnote

Douglas Whitaker is a doctoral candidate in Mathematics and Statistics Education at the University of Florida. His current research focus is on in-service statistics teachers' professional identities and assessing statistics.

Steven Foti is a doctoral student in Mathematics and Statistics Education at the University of Florida. His current research focus is on improving levels of statistical literacy in the community.

Tim Jacobbe is an Associate Professor of Mathematics and Statistics Education at the University of Florida. His research currently focuses on the assessment of students' and teachers' statistical literacy.

Introduction

The Levels of Conceptual Understanding in Statistics (LOCUS)¹ project aims to create valid and reliable assessments for conceptual understanding of statistics. The assessments are being developed in the spirit of the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) Pre-K-12 framework (Franklin et al. 2007) and the Common Core State Standards for Mathematics.

Statistics is a substantial component of the Common Core State Standards for Mathematics (CCSSM) (National Governors Association Center for Best Practices and Council of Chief State School Officers 2010) in grades 6-12. Even before the CCSSM's release, adoption, and implementation, however, statistics and data analysis were emphasized in documents such as the *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics 1989) and the *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics 2000). Building on these widely known standards, there is a trend for requiring both students and teachers to know more statistics.

While statistics in the K-12 grades is taught within mathematics courses, statistics is recognized as distinct from mathematics outside of the school curriculum. In higher education, it is common for departments of statistics to exist in their own right, separate from departments of mathematics. While there are many similarities between mathematics and statistics, there are substantial differences which can lead to an incomplete or incorrect view of statistics. Cobb and Moore (1997) argue that the discipline of statistics exists because of “the omnipresence of variability” and fundamentally differs from mathematics in the way that the context of a problem is treated: “In mathematics, context obscures structure... In data analysis, context provides meaning” (p. 803).

Due to the *omnipresence of variability* (Cobb and Moore 1997), conceptual understanding of statistics is difficult to assess. There is a strong tendency for assessments that primarily use multiple-choice items to assess computational and procedural understanding of statistics, if statistics is assessed at all (Gal and Garfield 1997). Instead, large-scale assessments may include items written to assess statistics, data analysis, and probability that actually capture mathematical rather than statistical thinking. de Lange (2007) identifies attention to the content domain as a key issue facing large-scale assessment. Large-scale assessments certainly have many constraints on their development, and these pressures may be manifested as construct underrepresentation, one of the major threats to validity

¹ National Science Foundation Grant No. DRL-1118168. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

(de Lange 2007; Messick 1994). Thus, many assessments which nominally assess statistics are apt to assess isolated topics in a narrow manner. Calls for improved assessment of statistics have been made (Shaughnessy 1992; Shaughnessy 2007), and progress has been made in answering them (delMas et al. 2007).

Many well-known assessments, which nominally assess statistics, either assess procedural understanding or do not assess statistics at all. A recently released multiple-choice item from NAEP 2013 at Grade 8 asked, “Rick has earned a total of 581 points on all of his science tests. His overall test average (mean) is 83. How many tests has Rick taken?” The content of this item was classified as *Data Analysis, Statistics, and Probability*, though it is arithmetic in nature. Items of this nature might better be described as *statistics-adjacent*. Although items like these address mathematical skills and may require recall of statistical facts, they do not assess statistical reasoning in a way that acknowledges the important role of variability and context. Increasing emphasis on statistics in the K-12 curriculum necessitates the development of valid, reliable assessments of conceptual understanding that take into account the distinction between mathematical and statistical reasoning.

The LOCUS Assessments

The LOCUS assessments were developed using a modified version of Mislevy and Riconscente’s (2006) evidence-centered design (ECD) process (for more detail, see Jacobbe et al. 2014). The assessments were written at two levels of statistical literacy: Beginning/Intermediate and Intermediate/Advanced. This choice was made because the GAISE framework (Franklin et al. 2007) hypothesizes a continuum along which students develop. To describe students’ understanding, this continuum is discretized as Levels A, B, and C. One illustration of the maturation in understanding across the levels is the focus on different types of variability at each level: at Level A, students focus on “variability within a group;” at Level B, students focus on covariability and “variability within a group and variability between groups;” at Level C, students focus on “variability in model fitting” (Franklin et al. 2007, 15).

The authors of the GAISE framework note, “Although these three levels may parallel grade levels, they are based on development in statistical literacy, not age” (Franklin et al. 2007, 13). A student in high school who has never been exposed to statistics will likely demonstrate understanding consistent with level A and have to progress through level B before being able to demonstrate understanding consistent with level C. With the varied student backgrounds, curricula, and standards used throughout the country, it is not reasonable to assume that students of any particular age or grade level operate statistically at level A, B, or C.

The GAISE Framework also describes the statistical problem-solving process as having four components: formulating questions, collecting data, analyzing data, and interpreting results. Together with Levels A, B, and C, this four-component statistical problem-solving process comprises the two-dimensional GAISE Framework. This two-dimensional framework was highly influential in the item-writing process for the LOCUS assessments.

The Beginning/Intermediate assessment corresponds with levels A and B of the GAISE framework, and the Intermediate/Advanced assessment corresponds with levels B and C. The pilot assessment consisted of four forms at the Beginning/Intermediate level and four forms at the Intermediate/Advanced level. The total piloted item pool consists of 48, 55, and 53 multiple-choice items at the beginning, intermediate and advanced levels, respectively, and 33 constructed-response items. Multiple-choice items were also categorized as belonging to one of the four statistical process standards identified in the GAISE framework.

Each Beginning/Intermediate form had 19 multiple-choice and 5 constructed-response items; each Intermediate/Advanced form had 21 multiple-choice and 5 constructed-response items. While some items were present on several forms, a key consideration was collecting student responses to all 189 items in the pool. Within each classroom, the four forms administered were spiraled at the student level. The assessments were given in either a single 90-minute period or two 45-minute periods, one period for the multiple-choice section and the other for the free-response section.

The Pilot Sample

The LOCUS assessments were piloted with 2075 students at the Beginning/Intermediate Level and 1249 students at the Intermediate/Advanced Level in April and May of 2013. Schools were selected for participation in this sample according to three criteria: a contact for the LOCUS project was in the area (for convenience of implementation), the school was in a high-performing district as measured by standardized assessment results, and it was in a state where statistics was taught prior to adoption of the CCSSM. The latter two criteria were important for avoiding floor effects with the assessment; including lower-performing schools or schools that did not include statistics instruction in their curriculum would likely yield data that would not be informative in the test-development process. Therefore, the results presented here reflect a best-case scenario for student performance on the LOCUS assessment. LOCUS assessments were administered in Arizona, Colorado, Florida, Georgia, New Jersey, and Ohio.

Of the 2075 students completing Beginning/Intermediate versions of the assessment, 2013 were in grades 6-8 and 62 were in grades 9-12; of the 1249 students completing Intermediate/Advanced versions of the assessment, all were in grades 9-12. Females and males accounted for 46% and 51% of all responses,

respectively; 3% of students did not respond. While this sample is not necessarily representative of all students in grades 6-12, the results from this sample provide an adequate snapshot of the statistical understanding of students that have seen some statistics in their curriculum prior to the implementation of the CCSSM.

Within a participating school, all students in eligible classrooms took the assessment. Data were analyzed only from students whose parents signed an informed-consent form; students were given a \$5 gift card to a major national retail store as an incentive for returning a completed informed-consent form – irrespective of whether they agreed to participate in the study or not. Recognizing that the students participating in the study were not taking the LOCUS assessments for a grade and therefore may not put forth a reasonable effort to complete the assessment, the researchers attempted to get ‘buy-in’ from the participating teachers who were given a small stipend for classroom supplies for returning their students’ completed informed-consent forms. We believe that this teacher buy-in – combined with instructions to the students indicating that their effort would be appreciated – resulted in many students putting forth good effort on the assessment. In some cases, it was clear to the researchers that a student did not put any effort into the assessment. Multiple-choice response sheets that did not demonstrate effort (e.g., the student used the bubbles to draw a picture) and constructed-response forms that were off topic in their entirety were not included in the analysis. Overall, we believe our incentive efforts and connections with excellent classroom teachers led to students putting forth a reasonable effort to perform their best.

One of the primary goals for the LOCUS project is to provide a characterization of grade 6-12 students’ current level of statistical understanding. This goal is particularly important considering the increasing role statistics has in these grades due to the CCSSM’s on-going implementation. Results from the LOCUS pilot assessment – administered in April and May of 2013 – are presented here to address this goal.

Results

The results from the LOCUS administration are presented in two sections: examples of items from the assessment are given first followed by a report of student performance on the multiple-choice items. Because of the richness and amount of data, student performance on constructed-response items is the subject of other manuscripts (e.g., Whitaker and Jacobbe 2014; Case and Jacobbe 2014). The overall performance of the students in the sample on items written to assess Levels A, B, and C, and the four components of the problem-solving process are given.

Examples of Items

The types of items that appear in the LOCUS assessment are intended to measure conceptual understanding of statistical concepts, different from those seen in previous assessments of statistics. Thus, we now present six sample multiple-choice items that appeared on the pilot assessments and are representative of the types of items appearing on the final forms of the assessments. The sample items presented were written to assess concepts at the beginning, intermediate, and advanced levels (GAISE levels A, B, and C) and are appropriate for students in grades 6-12. Sample items at each of the four statistical problem-solving processes are shown and, in some cases, illustrate the progression of depth of understanding required across the levels of the GAISE framework. The proportion of students choosing each response is given in parentheses, and the correct answer is bolded.

Students are trying to choose new school colors. Which is the most appropriate statistical question to be investigated for a study of school colors preference?

- (A) What colors are the colors of your main rival school? (4%)
 - (B) What colors were the old colors of your school? (7%)
 - (C) **Which colors are most popular in your school? (59%)**
 - (D) Which colors do you like best? (30%)
-

Figure 1. Sample Beginning Level (Level A) Item: Formulating Questions.

At the beginning level of the GAISE framework, as well as in the CCSSM (6.SP.1), students are expected to be able to identify a statistical question, which is a question with an answer based on data that vary, versus a question with a deterministic answer. The item given in Figure 1 requires students to be able to identify which question is statistical in nature. The questions in choices (A) and (B) are deterministic in nature; the main rival school and the students' school, have fixed colors that do not vary. The correct answer, choice (C), was selected by 59% of the students and is an appropriate statistical question that could be investigated in a study of school-colors preference. The question in choice (D) may also appear in a study of school-colors preference; however, it would most likely appear as the survey question during the collecting of data, and not the statistical question that is to be investigated.

Also at the intermediate level of the GAISE framework and in the CCSSM (7.SP.1), students are expected to begin understanding various sampling methods including ones that use random selection. The item given in Figure 2 assesses student understanding of both of these concepts. The student is required to select the question whose answer could be best approximated using a random sample of students at a school. The answer to choice (A) is a deterministic one; the number of students that attend the school is a fixed number that does not vary. It is also

unclear as to how one might use random sampling to answer the question. Choice (B) provides a question that *could* be answered using random sampling, but the population of interest is all the classes at the school. Therefore, it would make sense to take a random sample of classes at the school instead of students. The final distractor, choice (D), is a valid statistical question; however, it cannot be answered using a random sample of students from only one school.

For their final project, students in a math class are required to answer a question by collecting data about students at their school. For which of the following questions could a random sample of students provide the best approximate answer?

- (A) How many students attend the school? (16%)
- (B) How many hours does each class at the school meet per year? (11%)
- (C) **How many text messages do students at the school send per week? (58%)**
- (D) Do students at this school have higher test scores than students in other schools in the district? (15%)

Figure 2. Sample Intermediate Level (Level B) Item: Formulating Questions.

Rebecca wants to know how many books students in her school read over summer vacation. She attends a large school, and doesn't have time to ask every student. Which of the following would best allow her to make generalizations about all students in her school?

- (A) **Select 40 students from her school at random (84%)**
- (B) Select all of the students in her English class (3%)
- (C) Select the first 40 students that she sees after school (5%)
- (D) Select 40 of the students in the library at random (8%)

Figure 3. Sample Intermediate Level (Level B) Item: Collecting Data.

The collecting-data aspect of the statistical problem-solving process at intermediate level focuses heavily on random sampling, comparative experiments, and how random processes can be used in each. The item given in Figure 3 requires students to consider the design of a survey to ensure that the collected data can be used to make generalizations about the population of interest. Choice (A) would give the researcher the best chance of getting a representative sample of her school, and 84% of students were able to correctly identify this option as the best sampling method. Choice (B) would appeal to students who think that since the researcher is interested in books, they should choose students from their English class, who are likely to have read books over the summer. Choice (C) gives the illusion of randomness because there might not seem to be anything suspicious about the first 40 students that leave the school; however, this is not a proper random sampling technique. Finally, because choice (D) selects students from the library, the estimate of the number of books read is likely to be biased. Like the correct response, this distractor includes the buzzword “random.”

Lee wants to answer the question, “What proportion of sophomores at my high school plan to take a foreign language class during the next school year?” Which of the following methods would best allow Lee to answer his question?

- (A) Randomly select 50 students from the high school and ask them if they intend to take a foreign language class next year. (20%)
 - (B) Randomly select half of the foreign language teachers in the high school and ask them how many students are taking their classes this year. (8%)
 - (C) Randomly select half of the sophomores taking Spanish this year and ask them if they intend to take Spanish next year. (10%)
 - (D) Randomly select 40 sophomores from the high school and ask them if they intend to take a foreign language course next year. (62%)**
-

Figure 4. Sample Advanced Level (Level C) Item: Collecting Data.

The item given in Figure 4 is very similar to the previous item in that it addresses the use of random sampling in a survey design to help answer a question about a population of interest. At level C of the GAISE framework and in the CCSSM (S-IC.1), students are required not only to understand where random sampling is needed, but also understand that a random sample is required to attain a sample that is representative of the population of interest. Unlike the previous level B item, this item has the word “random” in all four response choices. Choice (A) is incorrect because the population of interest is limited to sophomores at the high school but this method does not include that limitation. Choice (B) is about trying to estimate how many students enrolled this year and does not address the proportion nor the students that are interested in enrolling next year. Choice (C) does not give all sophomores an equal chance of being selected in the sample. Finally, choice (D), which 62% of students chose as the answer, is a reasonable sampling method to answer the question of interest.

Joe and Tom attended a rally on a Thursday night to protest the removal of vending machines from their school. They both wondered what percentage of students from their school actually attended the rally. Joe decided to get an estimate the next day (Friday) by selecting a random sample of 25 students and asking each one if he or she was at the rally Thursday night. Six of the students (24 percent) said they attended the rally. On Monday, Tom selected a random sample of 50 students and asked each one the same question. Fifteen of the students (30 percent) said they had attended the rally.

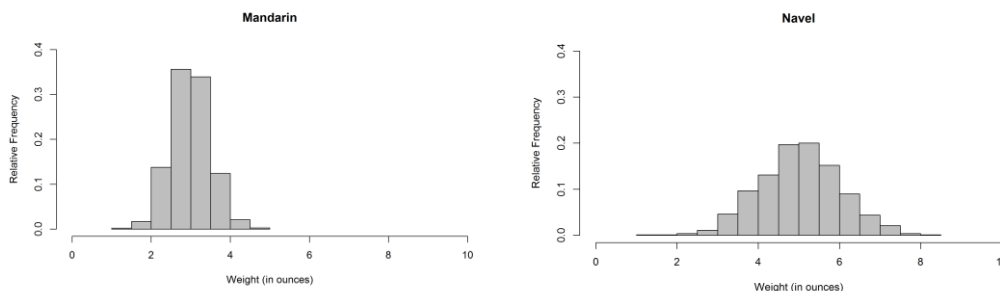
Which of the following is the most likely reason that Joe and Tom came up with different percentages?

- (A) A larger sample has more students and is more likely to have a higher percentage than the smaller sample. (36%)
 - (B) If they had both taken their samples on Friday the percentages would be the same. (3%)
 - (C) One of them did something wrong when selecting his random sample of students. (3%)
 - (D) The difference between the sample percentages could happen by chance with random sampling. (57%)**
-

Figure 5. Sample Intermediate Level (Level B) Item: Analyzing Data.

The LOCUS assessments recognize that analyzing data requires more than procedural knowledge of graphical and numerical summaries. The intermediate-level item given in Figure 5 assesses students' understanding of sampling variability, a concept that builds on understanding of other types of variability at level A and is developed more formally through descriptions of sampling distributions at level C. Choice (A) expresses the expectation of systematic differences in the sample percentages based on sample size. Choices (B) and (C) represent the expectation that every sample will produce the same summary statistic and attribute the difference to inconsistencies in sampling. The correct response, choice (D), acknowledges that different random samples from the same population are not likely to be identical.

A citrus farmer grows two types of oranges: mandarin and navel. The distributions of weight for each type of orange grown on his farm are shown in the histograms below.



The farmer will select a random sample of 100 mandarin oranges and a random sample of 100 navel oranges. Which sample is more likely to produce a sample mean within 1 ounce of the mean weight for its population?

- (A) **The sample of mandarin oranges because the weights of mandarin oranges are less variable than the weights of navel oranges. (55%)**
 - (B) The sample of navel oranges because the weights of navel oranges have a greater mean than the weights of mandarin oranges. (13%)
 - (C) Both samples are equally likely to produce a sample mean within 1 ounce because the sample sizes are the same. (18%)
 - (D) It is impossible to know which sample is more likely to produce a sample mean within 1 ounce without the actual data from the samples. (13%)
-

Figure 6. Sample Intermediate Level (Level B) Item: Interpreting Results.

To interpret results and make generalizations beyond the sample, students must be able to interpret the sampling variability in a summary statistic. The intermediate-level item given in Figure 6 assesses students' understanding of the variability of sample statistics over multiple samples and its relationship to variability in the population and to sample size. The concept assessed builds on understanding of natural variability within groups at level A and is critical for more formal significance testing and parameter estimation at level C. The correct

response, choice (A), recognizes that variation in sample statistics is related to variation in the population. Choice (B) incorrectly interprets sampling variability, because it confounds center and spread. Choice (C) assumes that sampling variability is related *only* to sample size and fails to acknowledge the connection to population variability. Choice (D) mistakes a random process for an unpredictable process and does not identify either sample statistic as less variable.

Overall Student Performance

The data from the pilot administration of the LOCUS assessments was analyzed using classical test theory (Crocker and Algina 1986); an analysis based on item response theory is planned for the operational administration. The reliability (as measured by stratified alpha) was between 0.70 and 0.72 for each of the four Beginning/Intermediate level forms and was 0.87 for each of the four Intermediate/Advanced level forms. The overall item difficulties (the proportion of correct responses) for the Beginning/Intermediate and Intermediate/Advanced forms were 0.52 and 0.53, respectively.

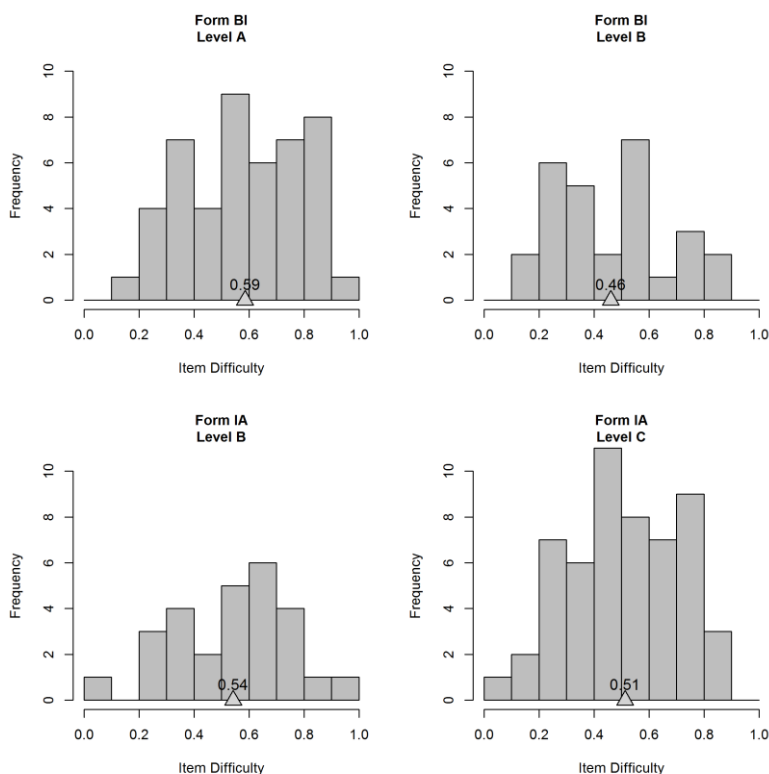


Figure 7. Frequency histograms of the item difficulties for items categorized as assessing Level A (48 items) and Level B (27 items) on the Beginning/Intermediate (BI) pilot forms and Level B (28 items) and Level C (53 items) on the Intermediate/Advanced (IA) pilot forms. A triangle indicates the mean on each histogram.

All items written to assess Level A concepts were piloted on the Beginning/Intermediate forms; all items written to assess Level C concepts were piloted on the Intermediate/Advanced forms. All forms included items written to assess Level B concepts. Frequency histograms for the item difficulties for items written to assess each level of the GAISE Framework are given in Figure 7. The average difficulties for Level A and B items on the Beginning/Intermediate forms were 0.59 and 0.46, respectively; the average difficulties for Level B and C items on the Intermediate/Advanced forms were 0.54 and 0.51, respectively.

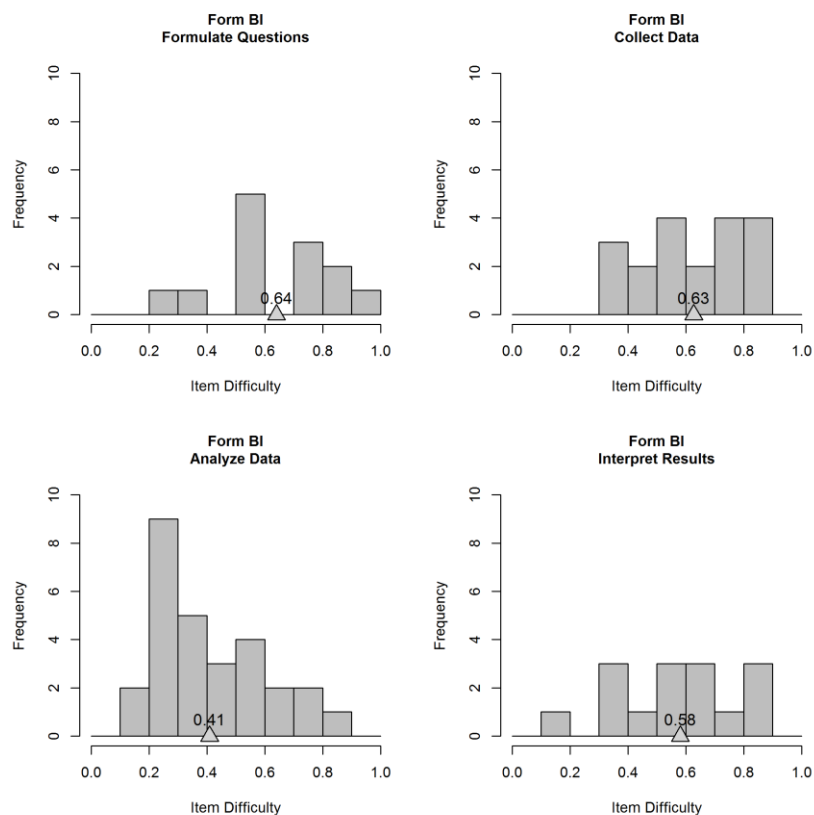


Figure 8. Frequency histograms of the item difficulties for items categorized as assessing Formulate Questions (13 items), Collect Data (19 items), Analyze Data (28 items), and Interpret Data (15 items) on the Beginning/Intermediate (BI) pilot forms. A triangle indicates the mean on each histogram.

Figures 8 and 9 are frequency histograms of the item difficulties for each component of the statistical problem-solving process on the Beginning/Intermediate and Intermediate/Advanced forms, respectively. On the Beginning/Intermediate forms, the average item difficulties for items assessing Formulate Questions, Collect Data, Analyze Data, and Interpret Results were

0.64, 0.63, 0.41, and 0.58, respectively. On the Intermediate/Advanced forms, the average item difficulties for items assessing Formulate Questions, Collect Data, Analyze Data, and Interpret Results were 0.68, 0.60, 0.51, and 0.38, respectively.

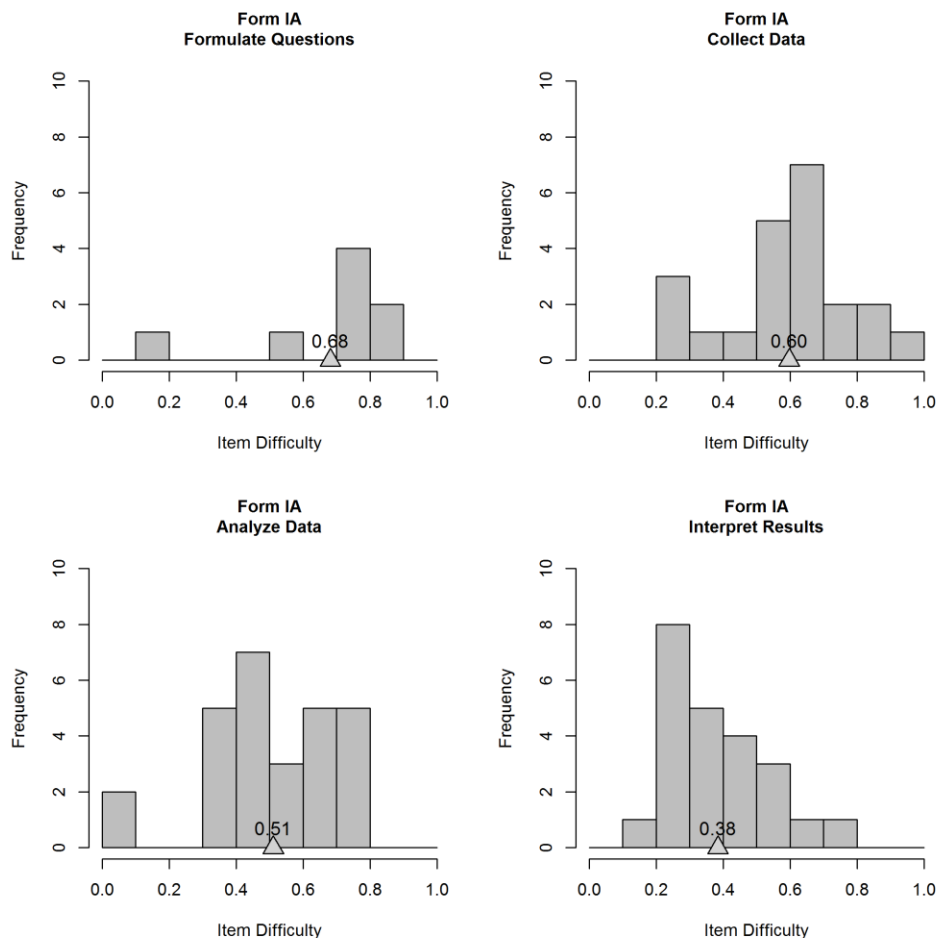


Figure 9. Frequency histograms of the item difficulties for items categorized as assessing Formulate Questions (8 items), Collect Data (22 items), Analyze Data (28 items), and Interpret Data (23 items) on the Intermediate/Advanced (IA) pilot forms. A triangle indicates the mean on each histogram.

The final test specifications for the LOCUS assessment call for each form to have approximately 40% of the items from the Formulate Questions and Collect Data components and 60% from the Analyze Data and Interpret Results components. Furthermore, the Beginning/Intermediate and Intermediate/Advanced forms will each have approximately 2/3 of the items drawn from the Beginning and Advanced pools, respectively, with the remaining 1/3 of the items drawn from the Intermediate pool; this was done to reflect the hypothesized

continuum of statistical literacy. These test specifications – along with the difficulty encountered in writing certain types of items – resulted in an item pool for the pilot administration that is not reflective of the final items to be used. Rather, additional items were written to ensure that the test specifications would be met with items whose performance in the pilot administration was acceptable. This has resulted in excess items for some levels of statistical literacy and components of the statistical problem-solving process which will be used for public dissemination and planned professional development materials.

Discussion and Conclusions

The results from the pilot administration of the LOCUS assessment provide a valuable snapshot of grades 6-12 students' understanding of statistics prior to the widespread implementation of the CCSSM. Students taking the Beginning/Intermediate version of the assessment – primarily students in grades 6-8 – tended to perform better on the Beginning Level (Level A) items than on the Intermediate Level (Level B) items. Students taking the Intermediate/Advanced version of the assessment – students in grades 9-12 – performed about the same on Intermediate Level (Level B) items as they did on Advanced Level (Level C) items. This is consistent with Watson and Moritz's (2000) finding that mastery of a lower level of statistical literacy is not a prerequisite for meaningful engagement at a higher level. Together with the hypothesized hierarchical levels of the GAISE Framework (Levels A, B, and C), this finding suggests that further research into the nature of the statistical literacy construct is necessary. Students in high school performed better on Level B items than students in middle school; while additional statistics instruction is the likely cause of some of the better performance observed with high school students, gains in statistical literacy due to age or students' maturation cannot be ruled out using these data.

Examining student performance on items assessing each of the four components of the statistical problem-solving process – Formulate Questions, Collect Data, Analyze Data, and Interpret Results – reveals other patterns. Both middle and high school students tended to perform better on questions assessing Formulate Questions and Collect Data; both groups also tended to perform relatively worse on the Analyze Data questions. While these conclusions are tentative pending a more thorough examination of the validity of the assessments for measuring these subconstructs, the nature of the questions categorized as Analyzing Data on the LOCUS assessments differs from a common perception of what these items look like: the questions assessing data analysis did not emphasize either computations or procedural understanding. One goal during the test development process was to write an assessment where the use of a particular calculator – or any tool at all – would not be a benefit. To that end, items were

written so that, where appropriate, both fractional and decimal versions of numbers were provided and ‘easy’ numbers (e.g., numbers divisible by 10) were used in computations. Discussions with the teachers who administered the LOCUS assessments and student responses to a survey item completed after taking the assessment corroborated our view that the use of a calculator is not a benefit for those taking the assessments. These results suggest that data analysis, conceived of as more than just a set of techniques and procedures, is still a challenging topic for students in grades 6-12, although progress is hopefully being made as students are exposed to more statistics in school.

The most-difficult material for students at the Intermediate and Advanced levels was Interpreting Results, with an average item difficulty of 0.38. At the higher levels of this component in the statistical problem-solving process, students are asked to interpret statistical results from relatively advanced topics such as margins of error, confidence intervals, and p -values (developed through simulations and randomization-based methods rather than based on asymptotic results). While p -values may not be frequently encountered in daily life, margins of error, confidence intervals, and other more-advanced topics are encountered. Most statistically literate citizens will not be called upon to produce statistical analyses; instead, most people interact with statistics as consumers, interpreting the results of others’ studies (Gould 2010; Wild 1994). These results suggest that many high school students are not yet capable of interpreting statistical results at a level suitable for statistical literacy.

Future Directions

The LOCUS assessments, which aim to measure conceptual rather than procedural understanding of statistics, were designed to be used by both researchers and teachers. Several online versions of the assessments are available online,² and paper-and-pencil versions are available from the authors. The research team is also working on a robust argument for the validity and reliability of the assessments in various contexts (e.g., with students and in-service teachers, among others) based on correlations with external measures and psychometric evidence, including Item Response Theory and Confirmatory Factor Analyses. In addition to providing valid, reliable instruments, the LOCUS team is developing materials to support professional development activities based on the copious amounts of student work collected during the project.

The CCSSM has the potential to be a boon for statistics: approximately 20% of the mathematics standards for grades 6-12 are statistical topics (Usiskin 2014). However, if statistics is not assessed authentically, it is likely that it will continue to be taught in a cursory manner at the end of the year – if at all (Madison 2015).

² <https://locus.statisticseducation.org/> Overview of LOCUS: Levels of Conceptual Understanding of Statistics (last accessed April 8, 2015)

The LOCUS assessments represent a vision for statistics assessments that measure more than computations and mathematical understanding. By aligning the assessments with the highly influential (Scheaffer and Jacobbe 2014) and American Statistical Association-endorsed GAISE Framework (Franklin et al. 2007), we seek to influence the writing of high-stakes, national assessments for students which claim to assess statistics. If statistics is assessed well, then we believe statistics instruction and statistics students will rise to meet the assessments.

The results of the LOCUS pilot assessments presented here provide a snapshot of what students with prior statistics instruction know. These data and results will serve as a baseline against which students' progress in understanding of statistics is measured as the CCSSM is implemented and assessed. Our society is not yet broadly statistically literate, but the LOCUS assessments may soon provide a way to measure changes in statistical literacy over time.

Acknowledgments

The LOCUS project would not have been possible without the entire project team. We would like to acknowledge the PI: Tim Jacobbe (University of Florida); the Co-PIs: Bob delMas (University of Minnesota), Jeff Haberstroh (Educational Testing Service), and Brad Hartlaub (Kenyon College); the Advisory Board Members: David Miller (University of Florida), Dick Scheaffer (University of Florida), J. Michael Shaughnessy (Portland State University), and Jane Watson (University of Tasmania); the Test Development Committee members: Christine Franklin (co-chair) (University of Georgia), Gary Kader (Appalachian State University), Mary Lindquist (Columbus State University), Jerry Moreno (John Carroll University), Roxy Peck (co-chair) (California Polytechnic State University), Mike Perry (Appalachian State University), Josh Tabor (Canyon del Oro High School); the External Evaluator: Jim Hammerman (TERC); and the Graduate Research Fellows: Catherine Case (University of Florida), Steven Foti (University of Florida), and Douglas Whitaker (University of Florida).

References

- Case, Catherine, and Tim Jacobbe. 2014. "Lessons from the LOCUS Assessments (Part 2): Center, Spread, and Informal Inference." *Statistics Teacher Network*, no. 84: 14–17.
- Cobb, George W., and David S. Moore. 1997. "Mathematics, Statistics, and Teaching." *The American Mathematical Monthly* 104 (9): 801–23.
<http://dx.doi.org/10.2307/2975286>
- Crocker, Linda M., and James Algina. 1986. *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart, and Winston.
- De Lange, Jan. 2007. "Large-Scale Assessment and Mathematics Education." In *Second Handbook of Research on Mathematics Teaching and Learning*, edited by Frank K. Lester, Jr., 1111–42. Charlotte, NC: Information Age Publishing, Inc.

- delMas, Robert, Joan Garfield, Ann Ooms, and Beth Chance. 2007. "Assessing Students' Conceptual Understanding after a First Course in Statistics." *Statistics Education Research Journal* 6 (2): 28–58.
- Franklin, Christine, Gary Kader, Denise Mewborn, Jerry Moreno, Roxy Peck, Mike Perry, and Richard Scheaffer. 2007. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*. August 2005. American Statistical Association. <http://www.amstat.org/education/gaise>
- Gal, Iddo, and Joan Garfield. 1997. "Curricular Goals and Assessment Challenges in Statistics Education." In *The Assessment Challenge in Statistics Education*, edited by Iddo Gal and J. B. Garfield, 1–13. Amsterdam ; Washington, DC: IOS Press.
- Gould, Robert. 2010. "Statistics and the Modern Student." *International Statistical Review* 78 (2): 297–315. <http://dx.doi.org/10.1111/j.1751-5823.2010.00117.x>
- Jacobbe, Tim, Catherine Case, Douglas Whitaker, and Steve Foti. 2014. "Establishing the Validity of the LOCUS Assessments through an Evidenced-Centered Design Approach." In *Sustainability in Statistics Education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.*, edited by Katie Makar and Rob Gould. Voorburg, The Netherlands: International Statistical Institute. http://icots.net/9/proceedings/pdfs/ICOTS9_7C2_JACOBBE.pdf
- Madison, Bernard. 2015. "Quantitative Literacy and the Common Core State Standards in Mathematics." *Numeracy* 8 (1): Article 11. <http://dx.doi.org/10.5038/1936-4660.8.1.11>
- Messick, Samuel. 1994. "The Interplay of Evidence and Consequences in the Validation of Performance Assessments." *Educational Researcher* 23 (2): 13–23. <http://dx.doi.org/10.3102/0013189X023002013>
- Mislevy, Robert J., and Michelle M. Riconscente. 2006. "Evidence-Centered Assessment Design." In *Handbook of Test Development*, edited by Steven M. Downing and Thomas M. Haladyna, 61–90. Mahwah, N.J.: Lawrence Erlbaum Associates.
- National Council of Teachers of Mathematics. 1989. *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author.
- . 2000. *Principles and Standards for School Mathematics*. Reston, VA: Author.
- National Governors Association Center for Best Practices, and Council of Chief State School Officers. 2010. *Common Core State Standards for Mathematics*. Washington, D.C.: Authors.
- Scheaffer, Richard L., and Tim Jacobbe. 2014. "Statistics Education in the K-12 Schools of the United States: A Brief History." *Journal of Statistics Education* 22 (2). <http://www.amstat.org/publications/jse/v22n2/scheaffer.pdf>
- Shaughnessy, J. Michael. 1992. "Research on Probability and Statistics: Reflection and Directions." In *Handbook of Research on Mathematics Teaching and Learning*, edited by D. Grouws, 465–94. Reston, VA: National Council of Teachers of Mathematics.
- . 2007. "Research on Statistics Learning and Reasoning." In *Second Handbook of Research on Mathematics Teaching and Learning*, edited by Frank K. Lester, Jr., 957–1010. Charlotte, NC: Information Age Publishing, Inc.

- Usiskin, Zalman. 2014. "On the Relationships Between Statistics and Other Subjects in the K-12 Curriculum." In *Sustainability in Statistics Education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.*, edited by Katie Makar and Rob Gould. Voorburg, The Netherlands: International Statistical Institute.
http://icots.info/9/proceedings/pdfs/ICOTS9_PL1_USISKIN.pdf
- Watson, Jane M., and Jonathan B. Moritz. 2000. "Development of Understanding of Sampling for Statistical Literacy." *The Journal of Mathematical Behavior* 19 (1): 109–36. [http://dx.doi.org/10.1016/S0732-3123\(00\)00039-0](http://dx.doi.org/10.1016/S0732-3123(00)00039-0)
- Whitaker, Douglas, and Tim Jacobbe. 2014. "Lessons from the LOCUS Assessments (Part 1): Comparing Groups." *Statistics Teacher Network*, no. 83: 13–15.
- Wild, Chris J. 1994. "Embracing the 'Wider View' of Statistics." *The American Statistician* 48 (2): 163–71. <http://dx.doi.org/10.2307/2684277>