

2011

Development of an Assessment of Quantitative Literacy for Miami University

Rose Marie Ward

Miami University - Oxford, wardrm1@miamioh.edu

Monica C. Schneider

Miami University - Oxford, schneimc@muohio.edu

James D. Kiper

Miami University - Oxford, kiperjd@muohio.edu

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



Part of the [Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

Ward, Rose Marie, Monica C. Schneider, and James D. Kiper. "Development of an Assessment of Quantitative Literacy for Miami University." *Numeracy* 4, Iss. 2 (2011): Article 4. DOI: <http://dx.doi.org/10.5038/1936-4660.4.2.4>

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

Development of an Assessment of Quantitative Literacy for Miami University

Abstract

Quantitative Literacy is a competence as important as general literacy; yet, while writing requirements are seemingly ubiquitous across the college curriculum, quantitative literacy requirements are not. The current project provides preliminary evidence of the reliability and validity of a quantitative literacy measure suitable for delivery online. A sample of 188 undergraduate students from Miami University, a midsize university in the midwestern U.S., participated in the current study. Scores on the measure, were inversely related to statistical/mathematical anxiety measures, directly related to subjective assessment of numeracy, and did not differ across gender or year in school. The resulting measure provides a reasonable tool and method of assessing quantitative literacy at a midsize university.

Keywords

Quantitative Literacy, assessment, college students

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

Rose Marie Ward is an Associate Professor of Health Promotion at Miami University. Her teaching and research interests examine college student health behavior change – specifically alcohol consumption and risky sexual behaviors. She is strong advocate for quantitative and statistical literacy.

Monica C. Schneider is an Assistant Professor of Political Science at Miami University. Her major research focuses on the study of stereotypes of female candidates in American politics. She has a particular interest in making sure that students gain enough quantitative literacy to effectively participate in a democracy.

James D. Kiper is Professor and Chair of the Department of Computer Science and Software Engineering at Miami University. His teaching and research interests lie in the areas of software engineering (particularly in design rationale) and in the scholarship of teaching and learning, especially in quantitative literacy and support for mobile learning.

Introduction

Quantitative Literacy (QL) broadly refers to one's ability "to apply simple mathematical methods to the solution of *real-world problems*" according to the Sons Report (Sons 1996) of the Subcommittee on Quantitative Literacy Requirements of the Mathematical Association (MAA) Committee on the Undergraduate Program in Mathematics (CUPM). While numeracy or quantitative reasoning is a key learning outcome for college graduates (Sons 1996; AAC&U 2011), it is unclear that every college graduate can apply mathematical methods to solve problems in everyday life.

Specifically, the Sons Report continues, a quantitative literate graduate should be able to:

1. "Interpret mathematical models, such as formulas, graphs, tables, and schematics and draw inferences from them.
2. Represent mathematical information symbolically, visually, numerically, and verbally.
3. Use arithmetical, algebraic, geometric, and statistical methods to solve problems.
4. Estimate and check answers to mathematical problems in order to determine reasonableness, identify alternatives, and select optimal results.
5. Recognize that mathematical and statistical methods have limits." (Sons 1996).

These two quotations provide a framework to discuss the level of QL in the United States.

It is estimated that approximately half the adults in the U.S. have very basic or lower quantitative skills (Kirsch et al. 2002; Kutner et al. 2006). Like verbal literacy, QL is positively related to important outcomes (for an exception see Butcher, McEwan, and Taylor 2010). For instance, there is a relationship between QL and understanding health outcomes. Fagerlin et al. (2007) found that those with self-reported high QL ability, compared with those with a low QL ability, have an easier time interpreting health-related data and thereby make more sound medical decisions. Rothman et al. (2006) relate numeracy (another name for QL) to understanding nutrition labels; those with lower numeracy had a more difficult time making interpretations necessary for sound nutritional decisions. In addition, patients with higher levels of numeracy were able to understand better how mammography reduces risk (Schwartz et al. 1997). Furthermore, QL skills are crucial for a higher probability of employment (Rivera-Batiz 1992) and advancement within a job. However, only 32% of surveyed employers, in one

survey, agree that college graduates are very well prepared in this area (Peter D. Hart Research Associates Inc. 2008). Finally, those with lower numeracy were more susceptible to framing effects (i.e., the propensity to choose positively over negatively worded outcomes even when the options are substantively equivalent) and more likely to be influenced by “competing, irrelevant affective considerations,” (Peters et al. 2006, p. 407) making them generally poorer decision makers.

In light of the importance of QL skills, organizations such as the AAC&U recommend that colleges and universities make QL skills and the assessment of those skills a priority (AAC&U 2011). Complementing the list of skills from the Sons Report (Sons 1996) and the VALUE rubric (AAC&U 2011), the current study sought to develop an assessment of QL appropriate for a midsize university.

Background

Quantitative Literacy: A Definition

While many readers are undoubtedly aware of the numerous QL definitions (e.g., Steen 2001; Madison and Steen 2008; Nelson et al. 2008; Steele and Kilic-Bahi 2008), a few key definitions are highlighted below. Common to all of the definitions is that QL is the ability to apply mathematical operations to real-world problems. For example, Kirsch and colleagues (1993, pp. 3–4) suggest that it is

the knowledge and skills required to apply arithmetic operations, either alone or sequentially, using numbers embedded in printed material (e.g., balancing a checkbook, completing an order form).”

Montori and Rothman (2005, p. 1070) further clarify this definition by stating that

the specific aspect of literacy that involves solving problems requiring understanding and use of quantitative information is sometimes called numeracy. Numeracy skills include understanding basic calculations, time and money, measurement, estimation, logic and performing multistep operations. Most importantly, numeracy also involves the ability to infer what mathematical concepts need to be applied when interpreting specific situations, and to use this information to problem solve.

The AAC&U (2011) and Sons report (1996) contend that quantitative reasoning is a habit of mind that is to be used in everyday life situations. In short, QL is one’s ability to use mathematics and statistics in everyday life.

Defining QL requires not only enumerating the skills that define it, but also distinguishing the concept from other, more familiar learning outcomes. In short, it is as important to specify what quantitative literacy *is not* as what it *is*. Lynn Steen, one of the founders of the QL movement, makes the point that quantitative literacy is not simply mathematics or statistics (Steen 2001, 2004). While

mathematics deals with abstract concepts, QL focuses on decision making and problem solving in *real-world situations*. As a result, someone could be highly proficient in advanced mathematics without possessing quantitative literacy.

Several institutions for higher education have sought to develop QL requirements, based on one or more of the QL definitions (Ward et al. 2010).¹ For example, at Vassar College, New York, students must take a QL course before their third year (Vassar College 2010). Hollins University, Virginia, has instituted QL in a variety of courses across the curriculum (Diefenderfer, Doan, and Salowey 2004). Hamilton College, New York, has a center for QL and a course requirement (Hamilton College 2010). Furthermore, the state of Washington created quantitative reasoning as an “incentive initiative” (Davidson and McKinney 2001). In addition, the State Council of Higher Education for Virginia (2006) names “Quantitative Reasoning” as one of the six core competencies for public institutions.

While QL by itself is important, it cannot be examined in isolation. Fear of math, numbers, or statistics is a common affliction that can inhibit students from performing well on exams in a statistics course (Keeley, Zayac, and Correia 2008). Moreover, efforts that enhance the “mathematical self-efficacy” of students also raise test scores (Maier and Curtin 2005). Increasing students’ comfort and confidence in using numerical concepts has been shown to be important for evaluating student performance in an individual class (Bos and Schneider 2009) and a university-wide QL program (Steen 2001, 2004). In sum, a successful QL program should leave students with a lasting command of using numbers in their lives, in addition to lowering student anxiety—or conversely improving their confidence—towards mathematical and statistical topics.

Quantitative Literacy: Assessment

Assessments of broad learning outcomes like QL can serve a variety of needs from an institutional perspective and have been recommended by the AAC&U (2011). Assessment can provide an institution with baseline data on how students are doing at the current time period. Institutions can use an assessment to place students in courses or allow them to place out of particular requirements. Assessments can be used to gauge the effectiveness of a particular intervention, or even to set measures of success for graduating seniors (Gardner 2006). In short, assessment can assist in the planning of QL curriculum and courses, improvement of teaching practices, and improvement of undergraduate education in general (AAC&U 2011).

¹ Lynn Arthur Steen from St. Olaf College developed a list of QL Programs in U.S. Colleges and Universities (Madison and Steen 2008, p. 10–13). Between Steen’s list and our own research, we found over 40 schools with either some type of QL requirement, center, or courses that explicitly mention quantitative literacy as a learning outcome.

The approach to the assessment of QL varies widely across institutions. Although the currently available assessments provide a variety of different options for assessing QL, many fit a narrow set of goals rather than addressing the many facets of QL. The assessment used by Dartmouth College, New Hampshire (Korey 2000), and Lawrence University, Wisconsin (Jordan and Haines 2006), asks students to rate their beliefs on four factors: (1) confidence in QL competency, (2) perceived utility of QL, (3) positive beliefs about QL, and (4) the level of interest in QL. This type of assessment incorporates beliefs about math and statistics, which is only a portion of many definitions of quantitative literacy. Colby-Sawyer College, New Hampshire (Steele and Kilic-Bahi 2010), developed the following three measures of quantitative literacy: (1) a basic skills test, (2) a QL skill test, and (3) a survey measuring the attributes of quantitative learning (i.e., self-confidence, anxiety, value, enjoyment, and motivation). In addition, Colby-Sawyer College created a rubric for evaluating senior projects on their use of QL. Donna Sundre and colleagues (Sundre 2008; Sundre and Thelk 2010) of James Madison University, Virginia, developed a 26-item multiple-choice assessment that focuses on two areas: “the ability to use graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomenon” and “the ability to discriminate between association and causation and identify the types of evidence used to establish causation” (Sundre 2008). While Taylor (2009) calls this assessment “deep”—it asks enough questions to accurately measure the concepts—she also cites that its weakness is that it is “narrow” and reflects only two important areas. In addition, Ellington and Havor (2006) at Virginia Commonwealth University have a test with questions on topics such as unit analysis, interpretation of charts and graphs, proportional reasoning, counting principles, general percentages, percentage increase or decrease, the use of mathematical formulas, averages, and exponential growth. Unfortunately, not all these assessments are publically available nor do they meet all of the major components of quantitative literacy defined above.

Quantitative Literacy Learning Outcomes

Prior to developing our QL assessment, we developed learning outcomes to guide the measurement development process. The Sons Report (1996) and the AAC&U VALUE rubric (2011) inspired the process. Based on these reports, there are five broad requirements (see Table 1). For example, students should be able to “interpret” and “represent” mathematical information. In addition, students should be able to use various mathematical methods to solve problems and they should be able to “estimate and check answers to mathematical problems in order to determine reasonableness” (Sons 1996). Finally, students should recognize the limitations of various mathematical and statistical models.

Table 1
Outcomes Generated by QL Faculty Learning Community and their Alignment to Learning Outcomes of the Sons Report (1996)

Sons Report Learning Outcome	Specific Outcomes
Interpret mathematical models such as formulas, graphs, tables, and schematics and draw inferences from them	<ul style="list-style-type: none"> • Understand interval and rank data, and sets (nominal data) • Interpret graphs and multiple visual displays of information and data • Write narratives interpreting quantitative data and their meaning • Understand the difference between correlation and causality
Represent mathematical information symbolically, visually, numerically, and verbally	<ul style="list-style-type: none"> • Communicate quantitative information in written or graphical forms • Produce basic numerical and graphical summaries of data • Formulate a basic translation of real-world problems into symbolic abstractions that can be manipulated
Use arithmetical, algebraic, geometric, and statistical methods to solve problems	<ul style="list-style-type: none"> • Use quantitative data to determine which of two assertions is best • Incorporate quantitative measures of uncertainty in understanding assertions, such as those found in popular media • Have strategies for making decisions in the face of uncertainty and incomplete data • Read and interpret algorithmic descriptions of a process for solving a problem
Estimate and check answers to mathematical problems in order to determine reasonableness, identify alternatives, and select optimal results	<ul style="list-style-type: none"> • Develop a greater respect for scientific studies and data produced from well-designed sample surveys and experiments • Understand the importance of measurement issues • Be aware of the presence or absence of evidence in support of a claim and know various ways to assess the validity of evidence
Recognize that mathematical and statistical methods have limits	<ul style="list-style-type: none"> • Understand the limits of quantitative data, and explain why some ideas are not amenable to quantification
Objectives Not Otherwise Specified	<ul style="list-style-type: none"> • Recognize and find quantitative information in their discipline • Have a more positive feeling about the use of quantitative information and be more favorably disposed towards using it

A community of Miami University faculty with an interest in quantitative issues was established by the university's Center for the Enhancement of Learning, Teaching, and University Assessment. A *Faculty Learning Community* (FLC) is "a cross-disciplinary faculty and staff group of six to fifteen members...who engage in an active, collaborative, yearlong program with a curriculum about enhancing teaching and learning with frequent seminars and activities that provide learning, development, the scholarship of teaching, and community building" Cox 2004, p. 8). As part of weekly meetings, the FLC members read articles on QL, attended a Project Kaleidoscope conference on QL at Carleton College,² and consulted with a team from Hollins College. This group developed 17 learning outcomes that linked to a variety of QL definitions

² http://serc.carleton.edu/quirk/pkal_workshop08/index.html

(Table 1). These learning outcomes complement the headings of the requirements from the Sons report (1996) and the AAC&U guidelines (2011). For example these outcomes include “be able to understand interval and rank data, and sets (nominal data)” and “be able to use quantitative data to determine which of two assertions is best” (Sons 1996).

Unique Challenges for a Midsize University

A number of colleges and universities have implemented QL initiatives. While some of these QL programs have been deployed and assessed, assessing quantitative literacy at a midsize university, like Miami University, poses unique challenges. The main branch of the campus used in the current study consists of approximately 15,000 undergraduate students. There are five different academic colleges/divisions: Arts and Science, Business, Education, Health and Society, Engineering and Applied Science, and Fine Arts (Miami University 2010). The university grants doctoral degrees yet has a strong undergraduate focus. In addition, the university currently has over 70 undergraduate areas of study or major. Given the size and breadth of programs available, implementing a university-wide assessment or even establishing core rubrics for each program at Miami University is a daunting task. Moreover, implementing rubrics for assessing QL portfolios (e.g., VALUE; AAC&U 2011) would likely be relegated into senior capstones (due to the universal nature of the course across the curriculum) thereby limiting the assessment of QL to only one stage of academic development.

Possibly in contrast to smaller institutions, midsize or larger universities need an assessment of QL that is not time- or resource-intensive. For example, at Carleton College, a liberal arts college in Minnesota, professors must evaluate individual student papers for quantitative concepts according to a created rubric as part of their assessment of QL (Grawe, Lutsky, and Tassava 2010). Colby-Sawyer College, New Hampshire, has QL goals in individual departments and assessment is conducted in first-year seminars and senior capstones (Steele and Kilic-Bahi 2008). These institutions are smaller than a midsize institution with enrollments of 2,009 and 1,119 undergraduate students respectively (America's Best Colleges 2011). To be practical for a midsize institution, a successful assessment cannot be implemented on a department or course level.

While the majority of the QL initiatives have been employed in smaller universities and colleges, James Madison University, Virginia, is an exception with its 17,000 undergraduates (America's Best Colleges 2011). In a resource-intensive manner, James Madison has an assessment day for students (The Center for Assessment & Research 2010) to assess the QL (and other competencies) of its students. While this method of assessment is effective for this particular university, the resources and institutional commitment are not practical for the

assessment of QL at most midsize or larger universities. To respond to this challenge, the current project suggests using minimal resources by allowing students to complete the QL assessment on their own time frame and offer only small monetary incentives for completion of the assessment.

Summary and Purpose of the Current Study

Given the importance of QL with respect to employment (i.e., Peter D. Hart Research Associates Inc. 2008; Rivera-Batiz 1992), making health care decisions (i.e., Fagerlin et al. 2007; Rothman et al. 2006; Schwartz et al. 1997) and sound decision making (Peters et al. 2006), assessment of college students' current levels of QL is crucial for institutional learning and development. Assessments of QL have been previously employed at colleges and universities in time- and labor-intensive manners. The purpose of the current project is to develop an assessment appropriate for a midsize university that captures a variety of aspects of QL and is not time- or resource-intensive to administer.

Miami University intends to use this assessment for several purposes. First, the assessment will be used to establish a baseline level of our students' QL skills. Second, the assessment could be used to make the case that our students need further development in their QL skills. Finally, the data can help Miami University gauge the level of QL among graduating seniors.

QL Measurement Development

Several steps were conducted to develop the quantitative literacy measure. Using the FLC learning objectives, potential items were evaluated for inclusion in the assessment. The goal was to map items onto each learning objective without having a disproportionate number of items representing any one objective. Potential items ($n = 100+$) were selected from a variety of online and published resources. With permission when required, items were selected from delMas (2002), the *Assessment Resource Tools for Improving Statistical Thinking* (ARTIST) (Garfield et al. 2006) Web site ($n = 9$), the Lipkus et al. (2001) numeracy scale ($n = 2$), the Medical Data Interpretation Test (Schwartz et al. 2005) ($n = 2$), and various Web sites ($n = 8$; BrainMass.com, AceTheDat.com, <http://math.ucdenver.edu/~wbriggs/qr>). The authors wrote additional items ($n = 20$) to increase the breadth and depth of the quantitative literacy assessment. The items the authors wrote assessed QL in "ordinary contexts" (e.g., calculation of air temperature, measurements in a cake mix). A large variety of contexts (e.g., baking, temperatures, visiting restaurants, tree growth) were selected to represent a wide selection of interests and to not over-represent any one academic background. Response options for the questions were both multiple choice (with the four options randomized) and fill in the blank.

After several potential items ($n = 52$) were selected, we held cognitive interviews—where a small number of students from the test population (described below) think aloud while reading the questions/directions, but without answering the questions—to examine the wording of the potential items. The three cognitive interview participants were one sophomore and two juniors. They had three different academic majors (psychology, exercise science, and dietetics). Items were revised based on these interviews.

Next, three university honors students (two juniors, one senior; dietetics, psychology/pre-med and zoology/creative writing majors) were given the revised QL assessment. They rated the assessment questions for difficulty, face validity (i.e., whether or not the item aligned with the proposed learning objective), and discussed with the primary author their experience of the assessment. Items that required an extensive amount of time (i.e., more than a few minutes) or greatly frustrated the students were evaluated, re-written, or removed from the assessment. After these pre-tests, 44 items of varying difficulties were selected for the data collection. The items for which we have permission to publish here are included in Appendix A (supplemental file).

Methods

Participants

A stratified random sample (across year in school) of student university email addresses ($n = 1,500$) was generated. Email invitations for the “Miami Reasoning Study” were sent during the fall semester (mid-September) and spring semester (mid-January). Email reminders were sent approximately one week after the initial invitation. The email invitation described the study and provided a link to the online questionnaire. All participants were offered compensation (i.e., \$5) for participation and were eligible for prizes.

If participants scored within the top 10% of the assessment, they were entered into a drawing for a digital music player (valued at \$300). Participants who scored in the top 20% of the assessment were entered into a drawing for four digital music players (each valued at \$50) and ten \$50 gift cards. All participants were eligible for a drawing of ten \$25 gift cards. Incentives were selected to increase the participants’ motivation to participate in the assessment and to do well on the assessment.

The final sample consisted of 188 participants (12.5% response rate). The sample was composed of 119 female students (63%); 55 students with senior standing (29.%); 64 students from higher income families (family income greater than \$100K) (34%), and 161 Caucasian students (86%). Participants reported an average age of 20.9 ($SD = 1.3$) and an average GPA of 3.10 ($SD = .54$). For the academic year when the data were collected, Miami University was composed of

female students (52.2%), students from higher income families (income greater than \$100K) (56%), and Caucasian students (84.3%) with an average age of 20. Approximately 34% of the first-year class was in the top 10% of their high school class (Miami University, 2010). Other than a slight oversampling of female students and an under-sampling of high-income students, the sample is very similar to the larger population of students from which it was drawn.

Measures

The QL assessment measure was examined with respect to two published measures. These published measures provide evidence of the validity of the newly developed QL assessment.

Statistics Anxiety Rating Scale. (STARS; Cruise et al. 1985; Baloglu 2002). The STARS utilizes 51 items on a five-point Likert scale to assess students' anxieties when enrolled in a statistical course or completing statistical analyses. Higher scores reflect more anxiety. The STARS measure has six subscales. The first 23 items measure statistics anxiety and are rated from "No Anxiety" to "Very Much Anxiety." The *Worth of Statistics* has 16 items and measures the perceived relevance of statistics (e.g., "Since I am by nature a subjective person, the objectivity of statistics is inappropriate for me."). The *Interpretation Anxiety* subscale has 11 items (e.g., "Interpreting the meaning of a table in a journal article."). The *Test and Class Anxiety* subscale has 8 items (e.g., "Studying for an examination in a statistics course."). The *Computation Self-Concept* reflects the participant's attitudes toward statistics and has seven items (e.g., "I have not had math for a long time. I know I'll have problems getting through statistics."). The *Fear of Asking for Help* subscale has four items (e.g., "Going to ask my statistics teacher for individual help with material I am having difficulty understanding."). The final scale of *Fear of Statistics Teachers* has five items (e.g., "Statistics teachers are so abstract they seem inhuman."). The final items measure dealing with statistics and are rated from "Strongly Disagree" to "Strongly Agree." Items were modified for the current project to include both statistics and mathematics. Even with the modification, subscale levels and internal consistencies were consistent with the published values for the subscales (Cronbach's alphas ranged from .79 to .95; see Table 2 for alphas, means, and standard deviations). Published means and standard deviations are available in Table 2.

Subjective Numeracy Scale. (SNS; Fagerlin et al. 2007; Zikmund-Fisher et al. 2007). The SNS assesses numeracy without utilizing mathematical calculations. It assesses the participant's perceived mathematical ability and preference for numerical data. It is an eight-item scale in which the overall average score and two subscales are used. Higher scores reflect a higher perception of one's own numerical ability and higher preference for numerical information. The average

overall score was 3.87 ($SD = .68$). For the *Ability* subscale, the average score was 3.82 ($SD = .90$) while the average score on the *Preference* subscale was 3.91 ($SD = .69$). The items, published means and standard deviations are presented in Table 3. The internal consistencies or Cronbach's alphas for the *Ability*, *Preference*, and total scale are .80, .84, and .66, respectively.

Table 2.

Relationship Between the STARS and the Miami QL Assessment

Subscale	Published <i>M(SD)</i>	<i>M(SD)</i>	Cronbach's Alpha	Correlation w/ Total score	Correlation w/ Time
Worth of Statistics (<i>the relevance of statistics</i>)	38.1 (14.0)	35.15 (14.59)	.95	-.37***	-.01
Interpretation Anxiety (<i>Anxiety experiences when trying to interpret statistical results</i>)	27.8 (8.4)	25.50 (7.62)	.89	-.20*	-.12
Test and Class Anxiety (<i>Anxiety experienced while taking exams or statistics courses</i>)	24.7 (7.7)	24.92 (7.23)	.90	-.42***	.02
Computation Self-Concept (<i>Person's attitudes toward statistics</i>)	15.9 (6.5)	13.37 (5.70)	.88	-.13	.01
Fear of Asking for Help (<i>anxiety experienced when a person attempts to ask for help regarding statistical problems</i>)	10.7 (3.4)	9.93 (3.79)	.85	-.29***	.01
Fear of Statistics Teachers (<i>Perceptions of statistics teachers</i>)	11.2 (4.2)	10.51 (4.03)	.79	-.27***	.01

Note: All items were modified to "statistics/mathematics"; Higher scores mean more anxiety; "Time" = the time the online survey was open on the participant's computer; *** $p < .001$; ** $p < .01$; * $p < .05$; published means and standard deviations are from Baloglu 2002.

The QL, STARS, and SNS measures were presented in two orders. Order 1 was STARS and SNS first and the QL questions second. Order 2 was QL questions first and STARS and SNS questions second.

Results

The internal consistency of the QL items was assessed using the Kuder-Richardson formula (K20) which is used in situations where the items are scored dichotomously (i.e., correct/incorrect) and the items have a variety of difficulty levels (see Traub 1994 for a discussion of reliability assessments). The QL items had a K20 of .83 or high internal consistency. The 44 items ranged in difficulty (i.e., 3.2% to 91.0% of the participants getting the items correct; Mean Difficulty of Items, 59.7%; Standard Deviation of Difficulty, 0.22%). These results met the

authors' expectations of a QL assessment with good internal consistency and varying levels of difficulty among the items. Having items that were considered difficult and easy keeps the participants interested and yet not overwhelmed.

Table 3.

Relationship Between the Subjective Numeracy Scale and the Miami QL Assessment

Item	Published <i>M(SD)</i>	<i>M (SD)</i>	Correlation w/ Total score	Correlation w/ Time
1. How good are you working with fractions?	3.67 (1.51)	3.56 (1.10)	.26***	.02
2. How good are you working with percentages?	3.92 (1.47)	3.73 (1.08)	.24**	-.02
3. How good are you at calculating a 15% tip?	4.20 (1.54)	3.95 (1.18)	.24**	-.05
4. How good are you at figuring out how much a shirt costs if it is 25% off?	4.58 (1.40)	4.05 (1.03)	.24**	.05
5. When reading the newspaper, how helpful do you find tables and graphs that are parts of the story?	3.83 (1.43)	3.94 (.83)	.15*	-.09
6. When people tell you the chance of something happening, do you prefer that you use words (it rarely happens) or numbers (there is a 1% chance)?	3.53 (1.82)	3.40 (1.27)	.12	.03
7. When you hear a weather forecast, do you prefer predictions using percentages (there will be a 20% chance of rain today) or predictions using only words (there is a small chance of rain today)?	3.06 (1.90)	1.93 (1.10)	.11	-.004
8. How often do you find numerical information to be useful?	4.16 (1.50)	4.21 (.78)	.13	-.02

Note: 5 point Likert Scale; For items 1-4 5 = "extremely good"; For item 5 5="extremely helpful"; For item 6 5="always prefer numbers"; For item 7 5="always prefer words"; For item 8 5="Very Often"; *** $p < .001$; ** $p < .01$; * $p < .05$

Initial tests examined the QL assessment with respect to the comparison validation measures (STARS and SNS) and the time spent on the assessment. It was expected that the score on the QL assessment would be negatively related to the anxiety measures (the STARS subscales) and positively related to the Subjective Numeracy Scale. In addition, due to the assessment being online and unsupervised, it was anticipated that those who spent more time would also have higher scores on the QL assessment. The total score (i.e., the number of QL items that the participants got correct) and completion time (amount of time the QL assessment was open on the participants' computers) were correlated with the STARS and the SNS using a Pearson's correlation. Five of the six STARS scales were inversely correlated with the total score (see Table 2). Consistent with previous research (Keeley, Zayac, and Correia 2008), as the total score increased the participant's rating of their statistical/mathematical anxiety decreased.

However, there was no correlation between completion time and the anxiety measure. Five of the eight items on the SNS were positively correlated with the total score (see Table 3). As total score increased, the participants' perception of their numeracy ability increased. Contrary to expectations, completion time and the items on the SNS were not significantly related. In addition, there was no relationship between the time the exam was open on the participant's computer and overall score, $r(186) = -.02, p = .75$.

Table 4.
Examination of the Miami QL Assessment Across Gender

Variable	Male <i>M (SD)</i> <i>n = 67</i>	Female <i>M (SD)</i> <i>n = 119</i>	<i>t</i> -test	Cohen's <i>d</i>
Miami QL total score	27.83 (9.08)	26.19 (10.00)	$t(182) = 1.10,$ $p = .27$.17
Time the assessment was open	4923.13 (26478.96)	6739.82 (28798.01)	$t(184) = -.43,$ $p = .67$	-.07
STARS				
Worth of Statistics	34.63 (16.38)	35.45 (13.60)	$t(159) = -.34,$ $p = .73$	-.05
Interpretation Anxiety	22.62 (7.54)	27.13 (7.24)	$t(163) = -3.80,$ $p < .001$	-.61
Test and Class Anxiety	23.85 (6.87)	25.37 (7.36)	$t(161) = -1.30,$ $p = .20$	-.21
Computation Self- Concept	13.26 (5.88)	13.36 (5.66)	$t(165) = -.10,$ $p = .92$	-.02
Fear of Asking for Help	9.66 (3.57)	10.01 (3.90)	$t(165) = -.58,$ $p = .56$	-.09
Fear of Statistics Teachers	10.90 (4.32)	10.15 (3.73)	$t(165) = 1.18,$ $p = .24$.19
SNS				
SNS total	3.99 (.64)	3.81 (.64)	$t(159) = 1.68,$ $p = .10$.28
SNS Ability	4.04 (.91)	3.72 (.85)	$t(166) = 2.26,$ $p = .03$.36
SNS Perceptions	3.92 (.64)	3.92 (.66)	$t(159) = -.06,$ $p = .95$	<.001

Additional analyses examined the relationship between demographic variables (i.e., gender, year in school, and GPA) and the QL assessment total score. The hope of the measurement development process was that the developed assessment would not differ across gender, would differ across year in school, and would be positively related to GPA. Consistent with the expectations of the current study, males and females did not differ on their total score or completion time (Table 4). Examining the sample for potential confounds, gender differences

were examined across the STARS and SNS. Males and females did not differ on the majority of the scales of the STARS and SNS. The two exceptions are that males reported lower statistical interpretation anxiety and higher perceived numerical ability than females (Table 4). In contrast to expectations, year in school (i.e., freshman, sophomore, junior, senior) did not differ across total score, completion time, the scales of the STARS, or the scales of the SNS (Table 5). However, freshmen were greatly under-represented in the sample. GPA was significantly related to the overall score, $r(181) = .27$, $p < .001$, but was not related to time spent on the exam, $r(183) = -.02$, $p = .84$.

Table 5.
Examination of the Miami QL Assessment Across Year in School

	Freshman <i>M (SD)</i> <i>n = 4</i>	Sophomore <i>M (SD)</i> <i>n = 44</i>	Junior <i>M (SD)</i> <i>n = 53</i>	Senior <i>M (SD)</i> <i>n = 55</i>	F test
QL total score	29.25 (9.29)	26.16 (10.51)	25.68 (9.26)	26.62 (10.64)	$F(3, 151) = .20$, $p = .90$
Time the assessment was open	188.68 (281.68)	10314.68 (37944.01)	3132.19 (16038.59)	763.41 (2052.06)	$F(3, 152) = 1.62$, $p = .19$
STARS					
Worth of Statistics	29.00 (11.79)	37.46 (16.06)	38.56 (13.79)	35.69 (13.75)	$F(3, 133) = .64$, $p = .59$
Interpretation Anxiety	24.33 (2.52)	25.10 (5.79)	27.34 (8.54)	24.91 (7.84)	$F(3, 135) = 1.07$, $p = .67$
Test and Class Anxiety	21.67 (4.16)	25.59 (7.16)	25.48 (7.27)	25.50 (7.05)	$F(3, 134) = .29$, $p = .83$
Computation Self-Concept	12.00 (3.61)	14.63 (6.18)	14.22 (5.77)	13.60 (5.17)	$F(3, 138) = .38$, $p = .77$
Fear of Asking for Help	12.33 (5.86)	8.76 (3.11)	10.62 (4.02)	10.23 (3.64)	$F(3, 137) = 2.50$, $p = .06$
Fear of Statistics Teachers	8.67 (.58)	10.76 (3.74)	10.29 (3.99)	10.83 (4.13)	$F(3, 138) = .42$, $p = .74$
SNS					
SNS total	4.25 (.43)	3.65 (.66)	3.84 (.65)	3.67 (.65)	$F(3, 134) = 1.37$, $p = .26$
SNS Ability	4.12 (.80)	3.62 (.74)	3.76 (.85)	3.79 (1.04)	$F(3, 139) = .87$, $p = .46$
SNS Perceptions	4.08 (.14)	3.69 (.74)	3.93 (.67)	3.92 (.62)	$F(3, 134) = 1.32$, $p = .27$

To examine the QL assessment further, each item on the assessment was examined. It was hypothesized that those with lower levels of anxiety (as measured by the STARS) and higher levels of subjective numeracy would score higher on the QL assessment items. Participants who answered the most difficult

QL item correctly had lower statistical/mathematical anxiety and greater perceived numeracy than those who answered it incorrectly. Participants who answered the easiest QL item wrong had lower overall scores and more anxiety than those who got it correct. There was no statistical difference across genders, ethnicities, income, and year in school across the most difficult and easiest questions. (See Table A2 in the appendix for the item analysis).

Several other aspects of the QL assessments can be noted. Eighty participants answered every question. Fill-in-the-blank or explain-your-reasoning questions were most frequently skipped. Only two participants attempted only one question. There were no significant differences between the two orders of the survey (STARS/SNS first, then QL assessment vs. QL assessment first, then STARS/SNS). The participants who attempted every question achieved higher scores.

Discussion

The goal of the current project was to develop a short, non-labor intensive assessment of QL suitable for our midsize university. The discussions of the recent literature during a QL Faculty Learning Community inspired the QL learning outcomes and the subsequent development of the QL assessment. The results from the current project will lead to the use the QL assessment to assist in the planning of QL curriculum, improve teaching practices, and strengthen the skills of our graduates by evaluating students' level of QL over time. This project represents the first step in the process.

The 44-item assessment of Qualitative Literacy provides a broad assessment of the construct as inspired by the QL FLC discussions, the Sons report (1996), and the AAC&U VALUE rubric (2011). The relatively efficient, internally consistent measure was easy to deliver via an unsupervised Web-based survey. In addition, the measure was related to statistical/mathematical anxiety and subjective numeracy in the predicted ways. Given that statistics/mathematics are not the same as QL (Sons 1996), it is not surprising that the correlations between the QL assessment and statistical/mathematical anxiety are moderate. However, the correlations are in the hypothesized direction suggesting that as QL increases, statistical/mathematical anxiety decreases. In addition, the QL assessment was related to the Subjective Numeracy scale. As scores on the QL measure increased, perceptions of one's numeracy abilities also increased.

The current study's lack of demographic differences on the assessment is consistent with research that suggests that QL cannot be determined by education level, intelligence, or outward appearance (Nelson et al. 2008). The lack of demographic differences stresses the equal importance of QL across genders, GPA, and year in school.

However, it was anticipated that participants who were further in their education process (i.e., juniors and seniors) would achieve higher scores on the QL assessment. A combination of factors likely led to our contrary result. First, the relatively low ($n = 4$) number of first-year students limits the current study's ability to generalize the findings. Based on power analysis estimates using the effect size from the current study, a sample of 274 individuals from each academic level is needed to detect differences in the data. Second, the first-year students who completed the assessment seemed to be academically motivated as seen in their high GPAs and high scores. Future assessments should seek to be sensitive to this selection bias. Next, Miami has not implemented a QL-infused curriculum. Finally, the timing of the assessment was not ideal for across-year comparisons. Due to a university policy, the assessment was not launched until mid-September, four weeks into the fall academic semester. Adjustment to college for the first-year students, and having semester obligations might have led to students who were more conscientious or academically motivated selecting to participate. Future research might examine the timing of the assessment with regards to the academic semester.

Midsize schools face many difficulties when attempting to assess the QL of their students. Generally, it is not practical for a midsize school to conduct a portfolio assessment of the students' QL proficiency. In addition, while some midsize schools have implemented an assessment day to supervise the students while they complete a QL assessment, this option is also resource-intensive. Online assessments permit a midsize school to request the assessment while allowing for the participant to complete it at his or her own pace. The manner of delivery of the assessment (i.e., online and unsupervised) strengthens the applicability of the measure. Participants who spent more time on the QL assessment did not score better, have more anxiety, or have a lower GPA. This manner of assessment does not require a separate "assessment day" or highly intensive scoring. In short, the characteristics of the current study's QL assessment allows for the implementation of the assessment at a midsize university.

Limitations

The current study, while providing an initial examination of a QL assessment, is not without limitations. The sample of students was fairly small and did not provide enough participants to compute split-sample comparisons. In addition, the sample was not diverse enough across race or ethnicity to compute any differences across the measure. Given that the current study was an initial examination of a QL assessment, future studies should seek to have a larger sample to replicate the findings and extend them across different ethnic groups. In addition, future examinations should seek equal representation across the

academic spectrum (i.e., freshman, sophomore, junior, senior). Another potential limitation is a selection bias. Once the potential participant viewed the content of the survey, he or she may have discontinued with the assessment due to mathematical/statistical anxiety. While the two orders of presentation of the materials (i.e., STARS/SNS then the QL assessment, and QL assessment then the STARS/SNS) did not cause any differences on overall score, time spent with the assessment open, or the anxiety measure, future studies might further explore this possibility.

Concluding Remarks

The QL measure that resulted from the current study is easy to administer at a midsize university. It may also be possible to implement the QL assessment at other sized universities that want to augment their current practices. The measure is internally consistent and performed as predicted against a measure of statistical/mathematical anxiety and a measure of subjective numeracy. Furthermore, assessing the QL of a university's students can help inform the development of QL initiatives; these initiatives can also push students one step closer to being quantitative literate adults. While one would presumably never be proud to say that they cannot read, there is wide social acceptability of dismissing quantitative concepts by saying, "I am not good with numbers."

Miami University intends to use this assessment for several purposes. First, the assessment will be used to establish a baseline level of our students' QL skills. Second, as a result of the QL FLC and the QL assessment process, the College of Arts and Science at Miami University adopted a QL requirement in their divisional standards. The QL assessment will be used to evaluate the effectiveness of the requirement over time. Finally, we will use the data to gauge the level of QL among the Miami University students. At the beginning of each fall semester, the QL assessment is being collected. With more quantitative literacy initiatives and assessments to measure and improve those initiatives, we hope to graduate seniors who feel confident with their numerical ability and thereby increase the number of adults with sufficient QL abilities.

Acknowledgments

Thank you to Provost Herbst and Provost Skillings for providing funding for the assessment of Quantitative Literacy. We express our deep appreciation to the members of the Miami University Quantitative Literacy Faculty Learning Community for the two years of conversations and lunches that inspired the current project.

References

- AAC&U. See Association of American Colleges and Universities.
 Association of American Colleges and Universities. 2011. "VALUE: Valid Assessment of Learning in Undergraduate Education."
<http://www.aacu.org/value/metarubrics.cfm> (April 18 2011).
- America's Best Colleges. 2011. *U.S. News & World Report*.
- Baloglu, Mustafa. 2002. Individual differences in statistics anxiety among college students. *Personality and Individual Differences* 34: 855–865.
[http://dx.doi.org/10.1016/S0191-8869\(02\)00076-4](http://dx.doi.org/10.1016/S0191-8869(02)00076-4)
- Bos, Angela L., and Monica Schneider. 2009. Stepping around the brick wall: Overcoming student obstacles in methods courses. *PS: Political Science and Politics* 42 (2): 375–383. <http://dx.doi.org/10.1017/S1049096509090519>
- Butcher, Kristin F., Patrick J. McEwan, and Corrine H. Taylor. 2010. The effect of quantitative skills training on college outcomes and peers. *Economics of Education Review* 29: 187–199. <http://dx.doi.org/10.1016/j.econedurev.2009.06.001>
- Cox, Milton D. 2004. Introduction to Faculty Learning Communities. *New Directions for Teaching and Learning* 97: 5–23. <http://dx.doi.org/10.1002/tl.129>
- Cruise, R. J., Cash, R. W., and Bolton, D. L. "Development and validation of an instrument to measure statistical anxiety." Paper presented at the annual meeting of the American Statistical Association Statistics Education Section. Las Vegas, Nevada, August 1985.
- Davidson, Marcia, and Gary McKinney. 2001. Quantitative reasoning: An overview. *Dialogue* 8.
- delMas, Robert C. 2002. Statistical literacy, reasoning, and learning: A commentary. *Journal of Statistics Education* 10 (3).
http://www.amstat.org/publications/jse/v10n3/delmas_discussion.html
- Diefenderfer, Caren L., Ruth Alden Doan, and Christina Salowey. 2004. The Quantitative Reasoning Program at Hollins University. *Peer Review* 6 (4): 13–15.
- Ellington, Aimee, and William Haver. 2006. "Contribution of a first year mathematics course to quantitative literacy." In *Current Practices in Quantitative Literacy*, ed. Rick Gillman. Washington, D.C.: The Mathematical Association of America.
- Fagerlin, Angela, Peter A. Ubel, Dylan M. Smith, and Brian J. Zikmund-Fisher. 2007. Making numbers matter: Present and future research in risk communication. *American Journal of Health Behavior* 31 (1): S47–S56.
- Gardner, John. 2006. *Assessment and Learning*. Thousand Oaks, CA: Sage Publications.
- Garfield, Joan, Robert DelMas, and Beth Chance. 2006. "Assessment Resource Tools for Improving Statistical Thinking." <https://app.gen.umn.edu/artist/index.html> (January 6, 2011).
- Grawe, Nathan D., Neil S. Lutsky, and Christopher J. Tassava. 2010. A Rubric for Assessing Quantitative Reasoning in Written Arguments. *Numeracy* 3 (1).
<http://dx.doi.org/10.5038/1936-4660.3.1.3>
- Hamilton College. 2010. "The Quantitative & Symbolic Reasoning Requirement."
<http://www.hamilton.edu/academics/resource/qlit/requirements.html> (October 29 2010).

- Jordan, Joy, and Beth Haines. 2006. The role of statistics educators in the quantitative literacy movement. *Journal of Statistics Education* 14 (2).
- Keeley, Jared, Ryan Zayac, and Christopher Correia. 2008. Curvilinear relationships between statistics anxiety and performance among undergraduate students: Evidence for optimal anxiety. *Statistics Education Research Journal* 7 (1): 4–15.
- Kirsch, Irwin S., Ann Jungeblut, Lynn Jenkins, and Andrew Kolstad. 1993. *Adult literacy in America*. Washington D.C.: U.S. Department of Education.
- . 2002. *Adult literacy in America: A first look at the findings of the National Adult Literacy Survey* (NCES Report No. 1993-275; 3rd Edition) Washington DC: US Department of Education, National Center for Education Statistics.
- Korey, Jane. 2000. "Dartmouth College Mathematics Across the Curriculum Evaluation Summary: Mathematics and Humanities Courses." <http://www.math.dartmouth.edu/~matc/Evaluation/humeval.pdf> (September 15 2010).
- Kutner, Mark, Greenberg, Elizabeth, Jin, Ying, and Paulsen, Christine. 2006. "The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy." <http://nces.ed.gov/pubsearch/pubinfo.asp?pubid=2006483> (December 16 2010).
- Lipkus, Isaac M. and Peters, E. 2009. Understanding the role of numeracy in health: Proposed theoretical framework and practical insights. *Health Education and Behavior* 36: 1065–1081. <http://dx.doi.org/10.1177/1090198109341533>
- Lipkus, Issac M., Greg Samsa, and Barbara K. Rimer. 2001. General performance on a numeracy scale among highly educated samples. *Medical Decision Making* 21 (1), 37–44. <http://dx.doi.org/10.1177/0272989X0102100105>
- Madison, Bernard L., and Lynn Arthur Steen. 2008. Evolution of numeracy and the National Numeracy Network. *Numeracy* 1 (1).
- Maier, Scott R., and Patricia A. Curtin. 2005. Self-efficacy theory: A prescriptive model for teaching research methods. *Journalism and Mass Communication Educator* 59 (4): 352–364.
- Miami University. 2010. "Quick Facts." <http://www.miami.muohio.edu/about-miami/quick-facts/index.html> (October 29 2010).
- Montori, Victor M., and Russell L. Rothman. 2005. Weakness in numbers: The challenge of numeracy in health care. *Journal of General Internal Medicine* 20 (11): 1071–1072. <http://dx.doi.org/10.1111/j.1525-1497.2005.051498.x>
- Nelson, Wendy, Valerie F. Reyna, Angela Fagerlin, Isaac Lipkus, and Ellen Peters. 2008. Clinical implications of numeracy: Theory and practice. *Annals of behavioral medicine* 35 (4): 261–274. <http://dx.doi.org/10.1007/s12160-008-9037-8>
- Peter D. Hart Research Associates Inc. 2008. *How should colleges assess and improve student learning? Employers views on the accountability challenge*. Washington D.C.
- Peters, Ellen, Daniel Viastfjall, Paul Slovic, C.K. Mertz, Ketti Mazzocco, and Stephan Dickert. 2006. Numeracy and decision making. *Psychological Science* 17 (5): 407–413. <http://dx.doi.org/10.1111/j.1467-9280.2006.01720.x>
- Rivera-Batiz, Francisco L. 1992. Quantitative literacy and the likelihood of employment among young adults in the United States. *The Journal of Human Resources* 27 (2): 313–328. <http://dx.doi.org/10.2307/145737>

- Rothman, Russell L., Ryan Housam, Hilary Weiss, Dianne Davis, Rebecca Gregory, Tebeb Gebretsadik, Ayumi Shintani, and Tom A. Elasy. 2006. Patient understanding of food labels: The role of literacy and numeracy. *American Journal of Preventive Medicine* 31 (5): 391–398. <http://dx.doi.org/10.1016/j.amepre.2006.07.025>
- Schwartz, Lisa, Steven Woloshin, William C. Black, and H. Gilbert Welch. 1997. The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine* 127 (11): 966–972.
- Schwartz, Lisa, Steven Woloshin, and H. Gilbert Welch. 2005. Can patients interpret health information? An assessment of Medical Data Interpretation Test. *Medical Decision Making* 25, 290–300. <http://dx.doi.org/10.1177/0272989X05276860>
- Sons, Linda 1996. *Quantitative reasoning for college graduates: A supplement to the Standards*. Washington D.C.: Mathematical Association of America. http://www.maa.org/past/ql/ql_part2.html (accessed January 6, 2011).
- State Council of Higher Education for Virginia. 2006. "Core Competencies." <http://research.schev.edu/topicpages.asp?t=26> (October 29 2010).
- Steele, Benjamin, and Semra Kilic-Bahi. 2008. Quantitative literacy across the curriculum: A case study. *Numeracy* 1 (2). <http://dx.doi.org/10.5038/1936-4660.1.2.3>
- . 2010. Quantitative literacy: Does it work? Evaluation of student outcomes at Colby-Sawyer College. *Numeracy* 3 (2). <http://dx.doi.org/10.5038/1936-4660.3.2.3>
- Steen, Lynn Arthur. 2001. The case for quantitative literacy." In *Mathematics and Democracy: the Case for Quantitative Literacy*, ed. Lynn Arthur Steen, 1–22. Woodrow Wilson National Fellowship Foundation.
- . 2004. Everything I needed to know about Averages ... I Learned in College. *Peer Review* 6 (5): 4–8.
- Sundre, Donna. 2008. *The Quantitative Reasoning Test, Version 9 (QR 9): Test Manual*. Harrisonburg, VA: The Center for Assessment & Research Studies.
- Sundre, Donna, and A.D. Thelk. 2010. Advancing assessment of quantitative and scientific reasoning. *Numeracy* 3 (2). <http://dx.doi.org/10.5038/1936-4660.3.2.2>
- Taylor, Corrine H. 2009. Assessing quantitative reasoning. *Numeracy* 2 (2). <http://dx.doi.org/10.5038/1936-4660.2.2.1>
- The Center for Assessment & Research. 2010. "Assessment Day." http://www.jmu.edu/assessment/JMUAssess/Aday_Overview.htm (November 3 2010).
- Traub, Ross E. 1994. *Reliability for the social sciences: Theory and application*. Thousand Oaks, CA: Sage Publications.
- Vassar College. 2010. "Academic Information." <http://deanoffreshmen.vassar.edu/academic/> (October 29 2010).
- Ward, Rose Marie, Schneider, Monica C., and Kiper, James D. "Development and validation of a quantitative literacy assessment." Presentation at the Lilly Conference, Oxford, OH, November 17-21, 2010.
- Zikmund-Fisher, Brian J., Smith, Dylan M., Ubel, Peter A., and Fagerlin, Angela. 2007. Validation of the Subjective Numeracy Scale: Effects of low numeracy on comprehension of risk communications and utility elicitation. *Medical Decision Making* 27: 663–671. <http://dx.doi.org/10.1177/0272989X07303824>