The Quantitative Reasoning for College Science (QuaRCS) Assessment 2: Demographic, Academic and Attitudinal Variables as Predictors of Quantitative Ability

Katherine Follette
Stanford University, kfollette@amherst.edu
Sanlyn Buxner
University of Arizona, buxner@email.arizona.edu
Erin Dokter
University of Arizona, edokter@email.arizona.edu
Donald McCarthy
University of Arizona, dwmccarthy@gmail.com
Beau Vezino
University of Arizona, beauvezino@email.arizona.edu
Laci Brock
University of Arizona, lsb311@email.arizona.edu
Edward Prather
University of Arizona, eprather@email.arizona.edu

The Quantitative Reasoning for College Science (QuaRCS) Assessment 2: Demographic, Academic and Attitudinal Variables as Predictors of Quantitative Ability

Abstract

In this article, we explore the ability of demographic and attitudinal variables to predict student scores on the Quantitative Reasoning for College Science (QuaRCS) Assessment. Variables measured by the assessment include: students' academic choices and plans, attitudes and perceptions regarding mathematics, self-reported effort level, and basic demographics such as age, race/ethnicity, gender and disability status. As in previously published numeracy studies, we find significant score deviations according to gender, race/ethnicity, and disability status; *however*, the effect size of these correlations pale in comparison to the effect size of affective/attitudinal variables on QuaRCS score. A large number of variables with significant effects on QuaRCS score make the data well-suited to dimension reduction, and Factor Analyses reveal that a majority of affective variables can be collapsed into three underlying factors, which we call numerical self-efficacy, numerical relevancy and academic maturity. These three composite variables alone account for 32.4% of the variance in QuaRCS score. Two additional affective variables - self-reported effort and calculator usage – add 15.9% to the regression model. Together, these five variables account for nearly half of the variance in QuaRCS score. In contrast, academic and basic demographic variables, account for only 0.3% and 0.1% of the remainder, respectively. Furthermore, most demographic variables (including race and gender) do not have a significant effect on the regression model once affective variables have been accounted for.

Keywords

quantitative literacy, science literacy, assessment

Creative Commons License

Cover Page Footnote

Kate Follette is a NASA Sagan Postdoctoral Fellow at the Kavli Institute for Particle Astrophysics and Cosmology at Stanford University. This work was begun as a graduate student at the University of Arizona, and was supported by an NSF Transforming Undergraduate Education in STEM grant (NSF-DUE # 114-0398) and an NSF Graduate Research Fellowship. She taught general education astronomy courses as an adjunct instructor at Pima Community College from 2009-2014, and will be an assistant professor at Amherst College beginning January 2017. Her science research focuses on finding and characterizing young extrasolar planets, as well as the disks of gas and dust from which they form.

Sanlyn Buxner is an assistant research professor in the Department of Teaching, Learning and Sociocultural Studies at the University of Arizona. She teaches introductory science and research methods courses and supports institutional assessment of graduate and undergraduate programs. Her research includes examining science literacy and quantitative literacy in undergraduate science students and studying the impact of research and industry experiences for K12 teachers on classroom practice and student outcomes.

Erin Dokter is an associate professor of practice in the Office of Instruction and Assessment at the University of Arizona, where she serves as the Coordinator for the Certificate in College Teaching program, and as adjunct faculty in the Agricultural Education Department. She has taught general education astronomy courses as adjunct faculty at San Diego State University, Grossmont College, Southwestern College, and Pima Community College. Her research and teaching interests revolve around educational development, cross-disciplinary learner-centered pedagogies, and active learning spaces.

Don McCarthy is a research astronomer and Distinguished Outreach Professor at The University of Arizona where he teaches science to undergraduates in the general education program and also astronomy courses for undergraduate astrophysics majors and minors. His research expertise in infrared astronomy focuses on the development and application of techniques for high angular resolution from large ground- and spaced-based telescopes with an emphasis on exoplanets and Solar System objects. He received the 2012 Education Prize from the American Astronomical Society.

Beau Vezino is a Ph.D student in the Department of Teaching, Learning and Sociocultural Studies at the University of Arizona. He teaches several undergraduate STEM education courses. His research includes examining the cognitive and affective impact of the inclusion of engineering in the K-12 classroom. Beau's work aims to improve the preparation of K-12 educators to teach engineering and STEM integration by developing of pre-service courses, in-service workshops, and curriculum. Laci Brock is a Ph.D student in the Department of Planetary Science and the Lunar and Planetary Laboratory at the University of Arizona. Her research and teaching efforts focus on assessing undergraduate students' knowledge of cosmological and geological time in order to develop and implement interdisciplinary, time-related curricula into introductory undergraduate science courses. She also conducts scientific research, using adaptive optics on large telescopes to observe brown dwarfs and model the evolution of their atmospheres.

Edward Prather is an Associate Professor in the Department of Astronomy and Steward Observatory at the University of Arizona. He serves as the founder and Executive Director of the NASA-funded Center for Astronomy Education. Ed's scholarship focuses on college-level STEM education research and faculty professional development.

Introduction

Support for the development of scientific and mathematical literacy in general education science courses has grown among educational researchers and STEM educators alike in recent years, as these skills are seen as vital to the growth of an educated citizenry (DeBoer 2000). College students themselves have shown concern over their own mathematical achievement, with approximately 28% of a recent sample of more than 140,000 college freshmen reporting that they expect to need remedial work in math (Eagan et al. 2015).

One approach to better comprehend college mathematics achievement has been to examine performance in mathematics courses. A complementary approach has explored college students' performance in other quantitative courses, such as general education science, in which students are expected to further develop their mathematical knowledge and skills. Regardless of study context, valid, reliable, and well-tested diagnostics are essential.

The Quantitative Reasoning for College Science (QuaRCS) assessment, an electronically-administered multiple choice instrument designed to measure college students' numerical skills and attitudes toward mathematics, was developed over the course of eight semesters between 2010 and 2015. The motivations for the development of the QuaRCS, the selection of specific quantitative skills for assessment, and development and validation of the quantitative questions on the assessment are described in the companion to this paper (Follette et al. 2015, hereafter Paper 1). Key results from Paper 1 that are relevant to this work include:

- The length of the survey (25 quantitative questions, 24 non-quantitative demographic and attitudinal questions) was shown to be free of the effects of test fatigue among students who self-report that they expended effort.
- Both paper and online formats were explored. It was determined that the benefits of online assessment (out-of-class administration, ease of data collection, freedom from time constraints) outweighed the risks (lack of control for calculator and resource usage, idling).
- Scores of experts and students enrolled in STEM major courses are significantly different ($p<0.001$) than those of general education science students on the assessment, suggesting that the instrument is capable of distinguishing varied levels of numeracy from novice to expert.

This paper (Paper 2) focuses on correlations between performance on the quantitative portion of the QuaRCS assessment and the demographic and affective variables probed by the 24 non-quantitative questions. These non-quantitative questions fall into three broad categories – basic demographics, academic demographics, and affective/attitudinal variables.

The paper also attempts to identify the underlying variables that drive differences in QuaRCS scores.

Following a review of the varied literature on demographic and attitudinal predictors of college students' quantitative literacy in the remainder of this section, the next section introduces the non-quantitative questions on the QuaRCS and describes how they were developed. The subsequent section describes the results from a variety of quantitative analyses of student responses to these questions, and the penultimate section is focused on interpretation of these results. In the final section, we summarize the main conclusions of our analyses.

Demographic Predictors of Quantitative Literacy in College Students

Mathematical performance among college students has been associated with various cognitive measures, such as academic preparation and prior learned knowledge (Taub, Benson, and Szente 2014), generalized skills such as fluid reasoning and reading comprehension (Kelly and Gaustad 2006; Sparks and Lovett 2009), and the length of time elapsed since a previous math course (Johnson 1996). However, there remain concerns about the role basic demographics (e.g., gender and ethnicity) play in mathematics achievement, as well as disability status. Further, there has been growing interest in examining attitudinal factors' roles in performance, specifically self-efficacy (i.e., "individuals' convictions that they can successfully perform given academic tasks at designated levels", ibid.), which has been correlated with effort, interest, and performance throughout college and beyond (Eccles 1994; Correll 2001; Bong and Skaakvik 2003).

Basic Demographics.
The demographic variables that have been most extensively studied in mathematics and science education are gender and race/ethnicity. While decades of research have revealed an achievement gap in the mathematical performance of students by gender and ethnicity, these gaps have decreased significantly in recent decades (Xie and Killewald 2012; Voyer and Voyer 2014). Still, large-scale studies have reported a small but significant "female advantage" in course achievement across all subjects, including mathematics (Bridgeman and Wendler 1991; Correll 2001; Voyer and Voyer 2014). A "male advantage" remains on performance of standardized tests of mathematics, though this has been consistently tied to persistent attitudinal and psychosocial factors, rather than ability (Bridgeman and Wendler 1991; Correll 2001; Voyer and Voyer 2014).

Despite actual performance, studies of mathematics self-efficacy among college women demonstrate a significant "confidence gap", even among STEM majors (Sax 1994; Xie and Killewald 2012; Sax et al. 2015). Moakler and Kim (2014) examined the relationship between confidence and demographic information in a national survey of 335,842 college freshman and found that even women and minorities who reported above average confidence in other domains still reported lower confidence in mathematics, and their mathematical confidence levels were nearly 10% lower than males and non-minorities.

Learning Disabilities.
Studies of the mathematical abilities of college students with documented learning disabilities have produced mixed results, particularly when using standardized achievement tests to assess skill level and knowledge. Sparks and Lovett's 2009 review of 108 studies reported that "[college students with learning disabilities'] mean scores on reading, spelling, mathematics, and written language measures were largely in the average range," with many scores above the median. Other studies report underperformance on mathematics tests among college students with diagnosed learning disabilities (McGlaughlin, Knoop, and Holliday 2005).

Among individuals with confirmed disabilities, underperformance on tests of quantitative literacy and mathematics have been correlated with memory, reasoning, and perceptual difficulties (McGlaughlin, Knoop, and Holliday 2005; Kelly and Gaustad 2006; Sparks and Lovett 2009), rather than mathematics knowledge. This complicates the ability to accurately gauge students' mathematical skill level. Given that an inestimable number of students with unidentified/undiagnosed disabilities exist in any study and our imperfect understanding of how a learning disability may influence achievement on mathematics diagnostic tests, caution is warranted in making comparisons between students reporting and not reporting a learning disability.

Foundations of Mathematics Self-Efficacy, Attitudes, and Anxiety in Undergraduates

Academic Preparation and Choices.
Collegiate learning is an entirely new environment for many students, one with increased cognitive demand. Many students enter college with limited quantitative reasoning abilities (Durrani and Tariq 2009) and lack the interdisciplinary mathematics preparation required to be successful in introductory science courses in particular (Matthews et al. 2009).

Mathematical experiences and preparation in secondary education are highly influential and are a significant driving force behind choice of major in college (Betz and Hackett 1983; Moakler and Kim 2014). Lower mathematical achievement deters students from pursuing STEM disciplines, and those who do choose STEM majors are disproportionately likely to be young, male, come from higher socioeconomic background, have a parent with a STEM degree, and have had sufficient prior preparation in science and mathematics (Wang 2013).

Math avoidance is commonly present when undergraduate students select their introductory coursework. Introductory Earth science courses, for example, have a history of being perceived as "easy" in the United States, and have acquired a reputation of requiring qualitative rather than quantitative skills (Manduca et al. 2008). Gilbert et al. (2012) surveyed 1057 students in introductory geoscience courses across seven universities and found 73% (*N*=765) enrolled to fulfill a degree requirement. Of students who had not yet declared a major, approximately 80% reported they were "not very likely" or "definitely not" planning on pursuing a major in one of the natural sciences (Gilbert et al. 2012).

Attitudes.

Students' beliefs about and attitudes toward using and learning mathematics are likely solidified long before they reach college. Wang (2012) found that a cascade of mathematical experiences from 7th to 10th grade influenced 12th grade mathematics achievement, which in turn influenced the decision to major in a STEM discipline. In other words, many college students have strong, previously established beliefs and opinions regarding the utility of mathematics.

This may be, at least in part, due to a disconnect between how primary- and secondary-level mathematics is *taught* in the United States and how mathematics is *used* in college and beyond. Research has revealed that K-12 teachers' mathematical knowledge is more often procedural than conceptual. In one study by Ma (1999), fewer than 10% of American mathematics teachers could create a story problem for the equation $1¾ \div ½$, whereas 90% of Chinese teachers could. Perhaps more alarmingly, only 43% of the American teachers solved this problem correctly (compared to 100% of Chinese teachers).

Uncontextualized mathematics teaching can influence student beliefs about the relevance, applicability, and role of mathematics in both college-level coursework and daily life (Matthews et al. 2009). Wang (2012) noted that transparency of teacher expectations and the existence of emotional and cognitive support contributed to increased student mathematical motivation throughout high school. This suggests that instructors of general education college science courses who provide their students with proper support and encourage the building of quantitative skills, regardless of a student's mathematical background and

performance, may be more likely to foster positive attitudes toward mathematics in introductory courses.

Pre-service elementary school teachers, who represent a notable population of students in introductory science courses, are particularly likely to express negative attitudes toward mathematics (Matthews and Seaman 2007). These students then go on to teach future generations at a critical time in development where students are still shaping their motivational academic beliefs.

Anxiety and Self-Efficacy.
Attitudes toward mathematics appear to be interrelated with a number of other factors, including anxiety and self-efficacy. Math anxiety is well-documented in college students and has even been shown to persist through adulthood in some cases. Various levels of severity exist, with some students expressing purely test-induced math anxiety, while others report anxiety in all math-related contexts. The literature discussing mathematics self-efficacy shows a direct relationship between low levels of self-efficacy and higher levels of mathematics anxiety (Matthews and Seaman 2007; Durrani and Tariq 2009), as well as correlations between high math anxiety and negative attitudes toward the development of numerical skills (Durrani and Tariq 2009; Moakler and Kim 2014). Students lacking confidence can begin to manifest negative attitudes toward mathematics in general. This hinders mathematical learning and performance (Durrani and Tariq 2009), leading students to exhibit characteristics of "math avoidance."

Successful completion of mathematical prerequisite courses has been correlated with lower math anxiety and higher student competence (Durrani and Tariq 2009), but simply passing a math course in high school does not guarantee sufficient grasp to apply numerical skills in quick-paced undergraduate science courses (Matthews et al. 2009). In a study of 569 undergraduate students, only 3.3% believed that mathematics was not important in science (Matthews et al. 2009). At the same time, students with weaker mathematical backgrounds who were interviewed about math-heavy science courses reported frustration with pacing. For these students, inadequate background led to difficulty in keeping up with the material and eventual disinterest in the course. This suggests that too much content in science courses, coupled with the mathematics they already fear, may lead students to feel overwhelmed and underprepared, further undermining their confidence.

Research on student attitudes toward statistics sheds additional light on the complexity of student attitudes toward mathematics. Results from Schau et al.'s (1995) Survey of Attitudes Towards Statistics (SATS) suggest that students may overestimate their mathematical competence prior to the start of a course. Upon then receiving grades lower than first expected in the course, students rate the value or relevance of the course material lower than they did initially (Pierce 2006).

These studies and domains illuminate the complexity of students' mathematical performance, particularly how attitudinal factors have a striking effect on achievement. The QuaRCS study assesses incoming undergraduate students' quantitative foundations—conceptually and attitudinally – and how they may change over the span of a course.

Study Design and Methods

The models developed in this paper are derived from data collected during three semesters of QuaRCS administration. A summary of the three datasets is provided in Table 1 below. A majority of the data were collected at a large research university in the Southwest, however we also include data from two community colleges (one Southwest, one South), one selective liberal arts college (East) and one additional large research university (Midwest).

In all cases, a student is considered to have "completed" the assessment when all 25 quantitative questions are answered. In cases where students completed the assessment more than once, only the first administration was retained in this analysis. All comparisons of pre-semester and post-semester performance, as well as analysis of instructors' attitudes, beliefs and teaching techniques, is left for future work in this series.

Table 1
Summary of Study Sample

Semester	Number of (non-duplicate) pre-semester completions	Number of courses	Range of completions per course	Number of institutions by type
Spring, 2015	1488	20	9-139	1 R1, 1CC
Fall, 2015	46	1	42	1 LA
Spring, 2016	1243	17	8-290	2 R1, 1CC
Total	**2777**	**38**	**9-290**	**2R1, 2CC, 1LA**

Notes: Summary statistics for the three administrations of the QuaRCS reported in this paper. R1 = Research 1, LA = liberal arts, CC = community college

The precise wording, answer choices, and classification as basic demographic (BD), academic demographic (AC), or affective (AF) for each question is provided in Table 2 in the order of appearance on the assessment. The table also specifies the percentage of respondents who elected each answer choice. All questions for which students were allowed to choose multiple answers have been separated into a series of binary variables (1=chose, 0=did not choose) for analysis.

Basic Demographic Questions

Students taking the QuaRCS are asked four fundamental demographic questions - their age, race/ethnicity, gender, and disability status. These questions appear at the very end of the assessment (questions 19-22 of 24) in order to minimize the effects of stereotype threat (Steele and Aronson 1995; Spencer, Steele, and Quinn 1999). They are included on the assessment because previous numeracy assessments, detailed in the Introduction section of this paper, have shown statistically significant deviations in test scores according to all four of these variables.

The disability status question in particular was developed carefully in consultation with the disability resource center at a large public university, and it allows students to specify separately whether they have a physical, cognitive, or learning disability. If they wish, students may specify the nature of their disability. They may also elect not to specify their disability status at all by choosing the "I prefer not to specify" option.

Race/ethnicity options are listed alphabetically – African American, Asian, Caucasian (non-Hispanic), Hispanic/Latino, Native American, and Other. Students electing the "other" option may choose to fill in their ethnicity. Here too, students may elect an "I prefer not to specify" option, and may also specify more than one race/ethnicity with which they identify.

The gender question included traditional binary responses of male/female in addition to a choice of 'other' to accommodate students who identify as non-binary. The number of students electing ages over 35 in the sample was very small, so we collapsed all age categories above this into a single category. Students under the age of 18 are also excluded from the sample due to human subjects requirements.

Table 2
Non-quantitative QuaRCS questions, in order of appearance on the assessment.

No.	Question	Answer Choices	N	%	Cat
1*	Overall, how difficult were the questions in this survey?	very easy	229	8.2	AF
		easy	603	21.7	
		moderate	1386	49.9	
		difficult	424	15.3	
		very difficult	130	4.7	
		no response	5	0.2	
2	In your everyday life, how frequently do you encounter situations similar to problems in this survey?	almost never	637	22.9	AF
		about once per year	412	14.8	
		about once per month	897	32.3	
		about once per week	599	21.6	
		daily	230	8.3	
		no response	2	0.1	
3	Which of the following best describes your calculator usage while taking this survey? I used a calculator to answer:	all or almost all of the questions	182	6.6	AF
		about 75% of the questions	416	15.0	
		about 50% of the questions	703	25.3	
		about 25% of the questions	642	23.1	
		I didn't use a calculator at all, or used it on only one or two questions	832	30.0	
		no response	2	0.1	
4	How frequently do you do calculations in your everyday life?	Never	162	5.8	AF
		Infrequently	671	24.2	
		Sometimes	1101	39.6	
		Frequently	841	30.3	
		no response	2	0.1	
5	How frequently do you encounter graphs and tables in your daily life?	Never	461	16.6	AF
		Infrequently	1158	41.7	
		Sometimes	846	30.5	
		Frequently	310	11.2	
		no response	2	0.1	
6*	Where would you put mathematics (including: doing calculations, reading graphs and tables, reasoning with numbers, etc.) on the following scales between two opposite adjectives (4 point scale)	Interesting (1) to Boring (4)			AF
		1	422	15.2	
		2	944	34.0	
		3	729	26.3	
		4	675	24.3	
		no response	7	0.3	
		Useful (1) to Useless (4)			
		1	1257	45.3	
		2	1053	37.9	
		3	343	12.4	
		4	115	4.1	
		no response	9	0.3	
		Easy (1) to Hard (4)			
		1	299	10.8	
		2	979	35.3	
		3	999	36.0	
		4	488	17.6	
		no response	12	0.4	
		Fun (1) to Scary (4)			
		1	229	8.2	
		2	1125	40.5	
		3	1041	37.5	
		4	367	13.2	
		no response	15	0.5	
7*	Rate the degree to which you agree with the following statement: "I feel confident using numbers in my non-math courses"	Strongly Agree	788	28.4	AF
		Agree	1390	50.1	
		Disagree	473	17.1	
		Strongly Disagree	121	4.4	
		no response	5	0.2	

Table 2 Continued

No.	Question	Answer Choices	N	%	Cat
8*	Rate the degree to which you agree with the following statement: "I feel confident using numbers in my everyday life"	Strongly Agree	907	32.7	AF
		Agree	1514	54.5	
		Disagree	291	10.5	
		Strongly Disagree	60	2.2	
		no response	5	0.2	
9*	Rate the degree to which you agree with the following statement: "Numerical skills are important to the understanding of science"	Strongly Agree	1260	45.4	AF
		Agree	1341	48.3	
		Disagree	139	5.0	
		Strongly Disagree	27	1.0	
		no response	10	0.4	
10*	Rate the degree to which you agree with the following statement: "Numerical skills are important to my everyday life"	Strongly Agree	801	28.8	AF
		Agree	1560	56.2	
		Disagree	354	12.7	
		Strongly Disagree	57	2.1	
		no response	5	0.2	
11*[a]	Rate the degree to which you agree with the following statement: "I am satisfied with my current level of numerical/mathematical skill"	Strongly Agree	250	9.0	AF
		Agree	603	21.7	
		Disagree	345	12.4	
		Strongly Disagree	90	3.2	
		no response	1489	53.6	
12	How long ago was your last math course?	I am taking a math course now	838	30.2	AC
		Within the last year	1076	38.7	
		Within the last two years	498	17.9	
		Within the last three years	194	7.0	
		More than three years ago	171	6.2	
		no response	0	--	
13	How many mathematics courses do you plan to take in college (including any you've already taken and are taking now)	none	256	9.2	AC
		one	668	24.1	
		two	914	32.9	
		three or more	936	33.7	
		no response	3	0.1	
14	How many science courses do you plan to take in college (including any you've already taken and are taking now)	none	158	9.2	AC
		one	395	24.1	
		two	983	32.9	
		three or more	1237	33.7	
		no response	3	0.1	
15	Why did you choose to take this course? Check all that apply.	It is a prerequisite for courses in my major	561	20.2	AF
		To fulfill a university general education requirement	1861	67.0	
		It sounded interesting	1609	57.9	
		It sounded easy	313	11.3	
		I heard the class was good	708	25.5	
		I heard the instructor was good	685	24.7	
		Other	150	5.4	
		no response	4	0.1	
16	I am a college . . .	freshman	1254	45.2	AC
		sophomore	842	30.3	
		junior	370	13.3	
		senior	272	9.8	
		other	34	1.2	
		no response	5	0.2	

Table 2 Continued

No.	Question	Answer Choices	N	%	Cat
17*	Please select your major or majors from the list below.	Humanities	160	5.8	AC
		Arts	191	6.9	
		Social Sciences	634	22.8	
		Education	153	5.5	
		Science	231	8.3	
		Engineering Mathematics or Computer Science	130	4.7	
		Business-related	716	25.8	
		Health-related	256	9.2	
		Trade-specific	107	3.9	
		Journalism	69	2.5	
		General Studies	75	2.7	
		Undecided	212	7.6	
		Other	136	4.9	
		no response	5	0.2	
18*	I chose (or will choose) my major because: Check all that apply.	I like the subject	2079	74.9	AF
		I feel that it will help me get a job I will enjoy after graduation	1692	60.9	
		I feel that it will help me get a well-paying job after graduation	1235	44.5	
		I am good at it	1338	48.2	
		I chose a major that would avoid math as much as possible	333	12.0	
		I chose a major that would avoid writing as much as possible	127	4.6	
		I'm not sure yet	212	7.6	
		Other	77	2.8	
		no response	10	0.4	
19	My age is . . .	under 18	0	--	BD
		18-25	2651	95.5	
		26-35	94	3.4	
		36-45	16	0.6	
		46-55	10	0.4	
		56+	6	0.2	
		no response	0	--	
20	My gender is . . .	Male	1248	44.9	BD
		Female	1515	54.6	
		Other	11	0.4	
		no response	3	.01	
21*	With which racial or ethnic group(s) do you most identify? Choose all that apply.	African American	157	5.7	BD
		Asian/Pacific Islander	340	12.2	
		Caucasian (non-Hispanic)	1659	59.7	
		Hispanic or Latino	665	23.9	
		Native American	73	2.6	
		Other	72	2.6	
		I prefer not to specify	78	2.8	
		no response	5	0.2	
22*	Have you ever been diagnosed with any of the following? Please select all that apply.	A physical disability	78	2.8	BD
		A cognitive disability	86	3.1	
		A learning disability	216	7.8	
		I prefer not to specify[b]	332	12.0	
		no response	2135	76.9	
23	Did you attend elementary, middle and high school entirely in the United States.	Yes	2490	89.7	AC
		No[c]	284	10.2	
		no response	3	0.1	

Table 2 Continued

No.	Question	Answer Choices	N	%	Cat
24	Knowing that this survey is being used for research to try to improve courses like yours and that your answer to this question will not be shared with your instructor, please honestly describe the amount of effort that you put into this survey.	I just clicked through and chose randomly to get the participation credit	152	5.5	AF
		I didn't try very hard	286	10.3	
		I tried for a while and then got bored	1035	37.3	
		I tried pretty hard	532	19.2	
		I tried my best on all or most of the questions	770	27.7	
		no response	2	0.1	
Calc.	Elapsed Time	Length of time taken to complete the instrument			NA
Calc.	Average Confidence Ranking	Average of the twenty five confidence rankings appearing after each			AF

Notes: Non-Quantitative QuaRCS questions and answer choices are listed here in order of their appearance on the assessment. The percentage of respondents for "choose all that apply" questions sum to greater than 100% across categories. This is because students are allowed to elect multiple options, so the percentages reflect the proportion of the 2777 respondents choosing each option. The final two questions in the table are calculated quantities, and not actual questions on the assessment.
* Question reverse-coded to be consistently worded from negative to positive for the analyses described in subsequent sections.
a Due to an oversight, this question did not appear in the Spring, 2015 version of the assessment
b After each of these, students may choose to specify the nature of their disability if they wish.
c If no: In what country did you attend elementary school?, In what country did you attend middle school?, In what country did you attend high school?

Academic Demographic Questions

Students completing the QuaRCS are also asked to specify their class year, major, most recent mathematics coursework, plans for future science and math coursework, and whether they were educated entirely in the US. The purpose of these questions is to gain a sense of their academic interests and preparation.

Attitude and Perception Questions

A number of affective questions probe student beliefs about mathematics, their view of the utility of numerical skills in science and in their daily lives, as well as their confidence and satisfaction with their current level of numerical ability. They are also asked to specify *why* they chose their major and *why* they chose to take the course in which the QuaRCS was administered. The final question on the assessment asks students to specify the amount of effort that they expended on the assessment, which we also consider an attitudinal measure.

After each quantitative question on the assessment, students are asked to rank their confidence in their response. These confidence rankings are averaged into a composite "average confidence ranking" variable, which we also consider to be affective.

The initial introduction to the assessment contains the sentence: "These questions were designed to be answerable without a calculator, but you are welcome to use one if you choose". After completing the quantitative questions, students are asked on what proportion they used a calculator to compute their answer. (See Table 2 for precise wording.) This question does not fit naturally into any of our three broad categories (basic demographic, academic demographic

or attitudinal), but is most closely related to student affect, and so we include it in this third group of variables in our analyses.

Particular attention was paid to the development of a question that asks students to rate "doing mathematics" on a four point scale between opposite affective adjectives: useful and useless, interesting and boring, easy and hard, and fun and scary. The choice of a four rather than five point scale was driven by a preliminary study in which we administered this question in paper form to a single, large, general education science course. Approximately half of the students (*N*=61) were provided with a five point scale, which allows for a "neutral" stance. The other half (*N*=55) were provided a four point scale that forced a choice between one extreme over the other. The difference between student responses on the two scales is striking (Fig. 1).
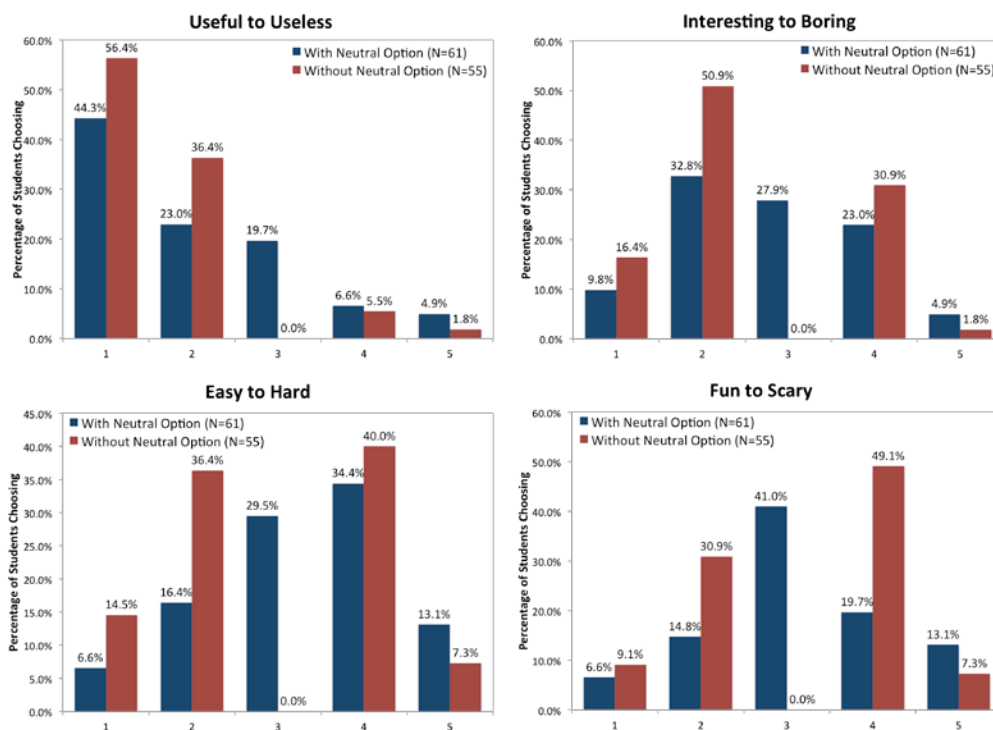

Figure 1. Normalized distribution of student ratings of "doing mathematics (including: doing calculations, reading graphs and tables, reasoning with numbers, etc.)" on either four (*N*=61, blue) or five (*N*=55, red) point scales between two opposite adjectives.

When forced to commit to one adjective or the other, it appears that students who would have selected a neutral option are more likely to shift toward the positive adjective (Useful, Interesting, Easy). The Fun vs. Scary scale, where students tended to move toward the negative option (Scary) when the neutral option is eliminated, is the exception to this rule.

Results

In the following section, we describe the construction of a model to predict students' scores on the QuaRCS assessment based on their responses to the non-quantitative questions on the assessment. In all subsequent discussions, "score" means the number of correctly answered quantitative questions, of which there are 25 (described in detail in Paper 1), and "variable" means a question on the assessment.

Mean Scores by Variable

As a first step in our analysis, we examined variations in score distributions among answer choices for all variables and identified binary and categorical variables where the 95% confidence intervals on the median score for a given group (indicated with notches in the box plots in Figures 2-4) do not overlap with the 95% confidence intervals on the median for the entire general education science student population (*N*=2777, indicated with horizontal lines). This is a very conservative means of identifying significant deviations, since inclusion of an individual group biases the overall population toward the group relative to a pure in vs. out-group comparison[1]. The full distributions are represented in Figures 2 through 4, and are grouped according to variable classification.

We find that all four fundamental demographic variables – age, race, gender and disability status contain categories that deviate from the overall distribution. In keeping with other numeracy assessment instruments, women, students of color, and the learning disabled underperform relative to other groups, and older students outperform younger students. An extensive literature, summarized in part in the first section of this paper, suggests that these differences are not reflective of innate differences in numerical ability, but rather other phenomena - namely that women are more susceptible to math anxiety (Betz 1978), people of color are underserved by the US educational system (Ogbu 1990; Raudenbush and Kasim 2009), and students with learning disabilities are particularly likely to struggle with applied mathematics (Chan and Dally 2001; Rousselle and Noël 2007). Additionally, despite being statistically significant, these effects are very small, as discussed in detail in the following sections of this paper.

A majority of academic variables also show at least one answer choice whose respondents' scores deviate significantly from the overall distribution, as shown

---

[1] Indeed, one-way ANOVA and *t*-tests for equivalence of means indicate many additional variables with significantly different mean QuaRCS scores. These more marginal effects can be seen in the more detailed analyses of later sections (notably Table 3) together with their effect sizes.

in Figure 3. Of particular note is that only STEM majors score significantly above the mean of the entire population, while only the means of humanities, education, and undecided majors lie significantly below it.
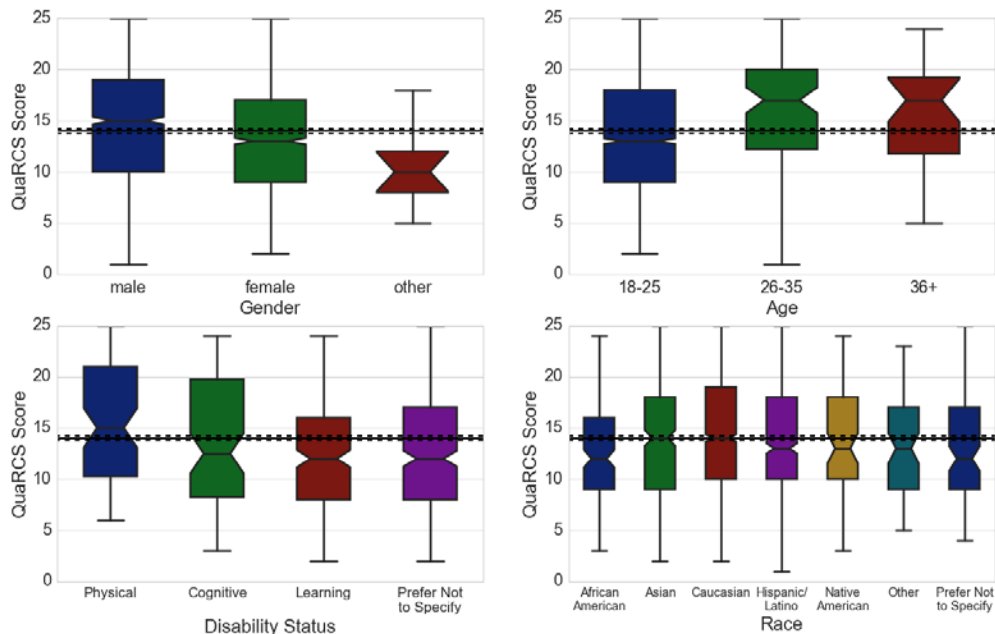

Figure 2. Average QuaRCS pre-semester scores (*y*-axis) according to responses to basic demographic questions (*x*-axes). In each panel, the median score for the entire pre-semester general education science population (*N*=2777) is shown as a solid horizontal line and 95% confidence intervals on the full population median are shown as horizontal dashed lines. The distribution of scores for each demographic question response is shown with filled, colored boxes indicating the 25 to 75% quartiles and whiskers extending across the full range of the score distribution for this group. The solid horizontal line in the middle of each box indicates the median of the distribution, and the notches show the 95% confidence intervals on this median Where the 95% confidence intervals for the medians of independent groups do not overlap, this supports a statistically significant difference between the group medians at the *p*<0.05 level.

The mean scores for QuaRCS affective variables show strong linear trends toward more positive attitudes (Fig. 4). Students answering in ways that suggest they have higher self-efficacy, perceive math as useful in their daily lives, and have devoted effort to the assessment score, on average, as many as 10 points higher than their counterparts with more negative views.

The distribution of confidence rankings is not easily shown as a box and whisker plot since student responses vary from question to question. Figure 5 represents the overall distribution of confidence rankings for all twenty five questions and all 2,777 students in the sample. The distribution of average confidence rankings for each student across the 25 individual questions is overplotted.
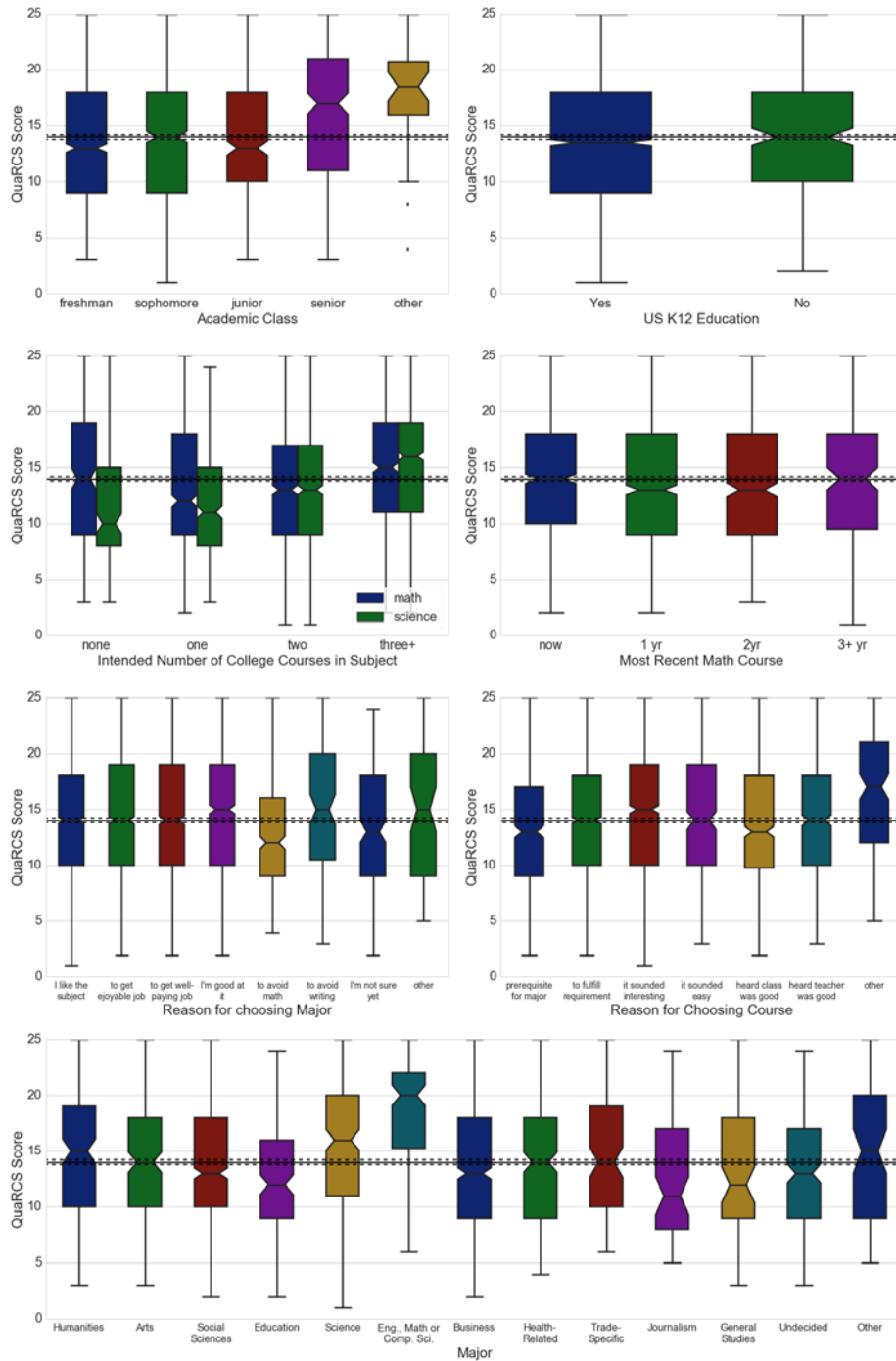
Figure 3. Average QuaRCS scores (*y*-axis) according to responses to various academic demographics (*x*-axes). Plot symbols, colors, and lines are described in detail in the caption to Figure 2.
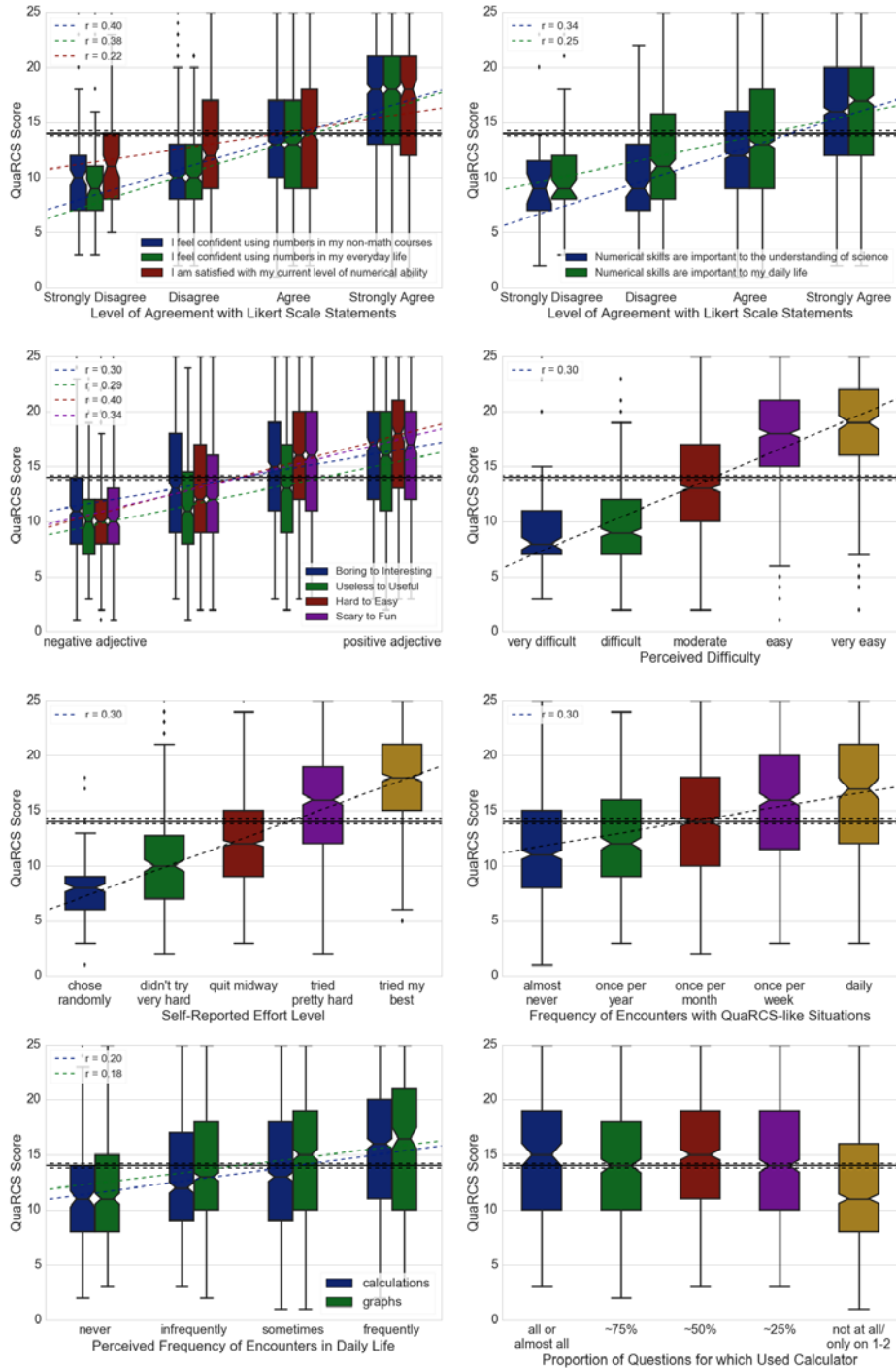
Figure 4. Average QuaRCS scores (*y*-axis) according to responses to attitudinal questions (*x*-axes). Where appropriate, regression lines are shown with *R* values indicated in the legend. Several panels show related questions in the same chart, with symbols described in their legends.
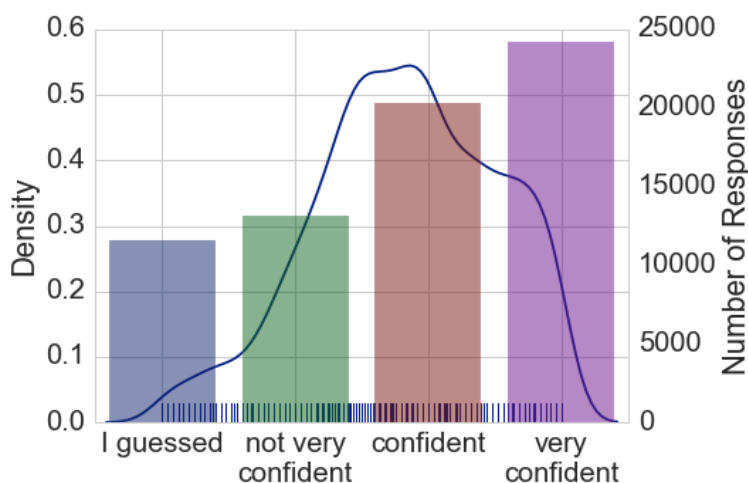
Figure 5. Distribution of responses to the confidence question: "How confident are you in the answer you just specified". The bar plot shows the distribution of the 69,327 responses given by each student after each quantitative question on the assessment. The line shows the average kernel density for a histogram of *average* confidence ranking for each student. The ticks on the bottom of the x-axis indicate densities of individual average values for the 2,777 students in the sample.

Effect Sizes

As Figures 2 through 4 suggest, the responses to many of the 24 supplemental questions on the QuaRCS are strongly associated with students' QuaRCs scores, however a small number are not, and in many cases for "choose all that apply" questions, only particular answer choices show significant relationships with score. In the analyses that follow, questions for which students can elect multiple answers are treated as a series of separate binary variables, making a total of sixty five variables. For all variables showing significant ($N < 0.05$) correlations with QuaRCS score, Pearson's correlation coefficients are listed as "effect size." Table 3 summarizes these significant correlations in order of the absolute value of their effect size.

The nineteen variables that are *not* significantly correlated with QuaRCS score are also worthy of note, and are listed in Table 4. In some cases, the lack of significant correlation may simply result from an inadequate sample size, but with a population of nearly 3000, such effects will be necessarily small. Of particular note is the lack of correlation between score and the length of time since a student's most recent math course, as well as the amount of time that elapsed between a student's starting and finishing the assessment.

Table 3
Variables Related to QuaRCS Score

No.	Question	Sig. (2-tailed)	Effect Size (*R*)	Cat.
Calc.	Average confidence ranking	<0.001	0.723	AF
24	Please honestly describe the amount of effort that you put into this survey.	<0.001	0.571	AF
1	Overall, how difficult were the questions in this survey?	<0.001	0.538	AF
7	"I feel confident using numbers in my non-math courses"	<0.001	0.405	AF
6c	Math is: Easy to Hard	<0.001	0.397	AF
8	"I feel confident using numbers in my everyday life"	<0.001	0.376	AF
9	"Numerical skills are important to the understanding of science"	<0.001	0.339	AF
6d	Math is: Fun to Scary	<0.001	0.337	AF
10	"Numerical skills are important to my daily life"	<0.001	0.250	AF
6a	Math is: Interesting to Boring	<0.001	0.298	AF
6b	Math is: Useful to Useless	<0.001	0.291	AF
2	In your everyday life, how frequently do you encounter situations similar to problems in this survey?	<0.001	0.282	AF
14	How many science courses do you plan to take in college?	<0.001	0.249	AC
11	I am satisfied with my current level of numerical ability	<0.001	0.217	AF
4	How frequently do you do calculations in your everyday life?	<0.001	0.202	AF
5	How frequently do you encounter graphs and tables in your daily life?	<0.001	0.185	AF
15c	I chose this course because: It sounded interesting	<0.001	0.173	AF
18d	I chose my major because: I am good at it	<0.001	0.176	AF
17f	Major: Engineering, Mathematics or Computer Science	<0.001	0.188	AC
3	Which of the following best describes your calculator usage while taking this survey?	<0.001	-0.136	AF
18b	I chose my major because: I feel that it will help me get a job I will enjoy after graduation	<0.001	0.128	AF
15g	I chose this course because: Other	<0.001	0.128	AF
16	Class year	<0.001	0.125	AC
20	My gender is . . .	<0.001	-0.132	BD
21c	Race/Ethnicity: Caucasian (non-Hispanic)	<0.001	0.108	BD
22d	Disability status: I prefer not to specify	<0.001	-0.099	BD
18e	I chose my major because: I chose a major that would avoid math as much as possible	<0.001	-0.093	AF
18c	I chose my major because: I feel that it will help me get a well-paying job after graduation	<0.001	0.091	AF
13	How many mathematics courses do you plan to take in college?	<0.001	0.091	AC
17e	Major: Science	<0.001	0.090	AC
18a	I chose my major because: I like the subject	<0.001	0.089	AF
22c	Disability status: learning disability	<0.001	-0.076	BD
19	My age is . . .	<0.001	0.075	BD
15a	I chose this course because: It is a prerequisite for courses in my major	<0.001	-0.073	AC
21a	Race/Ethnicity: African American	<0.001	-0.067	BD
17d	Major: Education	0.005	-0.054	AC
17l	Major: Undecided	0.008	-0.051	AC
17a	Major: Humanities	0.020	0.044	AC
21d	Race: Hispanic or Latino	0.032	-0.041	BD
18g	I chose my major because: I'm not sure yet	0.036	-0.040	AF
21g	Race: I prefer not to specify	0.049	-0.037	BD
15d	I chose this course because: It sounded easy	0.044	0.019	AF
18f	I chose my major because: to avoid writing as much as possible	0.005	0.053	AF
22a	Disability status: physical disability	0.024	0.043	BD

Notes: QuaRCS variables significantly associated with score are listed here in order of absolute value of effect size.
Question wordings in the table are, in most cases, more concisely worded than their actual appearance on the assessment.
The full phrasing can be found in Table 2. Question classifications as basic demographic (BD), academic demographic
(AC) or affective (AF) are also shown.

Table 4
Variables Unrelated to QuaRCS Score

No.		Question	Sig. (2-tailed)	Effect Size (*R*)	Cat.
12		How long ago was your last math course?	0.473	N/A	AC
15		Why did you choose to take this course? Check all that apply.			
	b	To fulfill a university general education requirement	0.616	N/A	AC
	e	I heard the class was good	0.599	N/A	AF
	f	I heard the instructor was good	0.678	N/A	AF
17		Please select your major or majors from the list below.			
	b	Arts	0.353	N/A	AC
	c	Social Science	0.272	N/A	AC
	g	Business-Related	0.523	N/A	AC
	h	Health-Related	0.435	N/A	AC
	i	Trade-Specific	0.162	N/A	AC
	j	Journalism	0.058	N/A	AC
	k	General Studies	0.163	N/A	AC
	m	Other	0.699	N/A	AC
18		I chose (or will choose) my major because: Check all that apply.			
	h	Other	0.254	N/A	AF
21		With which racial or ethnic group(s) do you most identify? Choose all that apply.			
	b	Asian/Pacific Islander	0.904	N/A	BD
	e	Native American	0.710	N/A	BD
	f	Other	0.373	N/A	BD
22		Have you ever been diagnosed with any of the following? Please select all that apply.			
	b	a cognitive disability	0.752	N/A	BD
23		Did you attend elementary, middle and high school entirely in the United States.	0.799	N/A	BD
Calc.		Elapsed Time	0.433	N/A	N/A

Notes: QuaRCS variables not significantly associated with score are listed here. In all cases, since the correlations are not significant (*p*>0.05), effect sizes are not given. Questions are categorized as basic demographic (BD), academic demographic (AC) or affective (AF).

Factor Analysis

Even if we include only those variables with significant effects on QuaRCS score in further analyses, we are still left with far too many (*N*=45) for a simple model. As 26 of the 45 significant variables were classified as affective (including eighteen of the twenty variables with the largest correlation coefficients), we engage in dimension reduction of these affective variables before computing a regression model for our data. Indeed, we fully expect that the affective questions on the QuaRCS probe a smaller number of underlying latent variables that cannot be measured directly, and dimension reduction is designed to identify these underlying factors. By combining questions into composite variables, we reduce the dimensionality of our data and extract underlying factors that are more robust and descriptive than individual variables alone.

We first completed an Exploratory Factor Analysis (EFA) on the Spring, 2015 data sample alone (*N*=1488) in order to build an initial model of composite variables. [2] We then cross-validated the structure of this model using Confirmatory Factor Analysis on the remaining sample (Fall 2015 and Spring 2016, *N*=1289). Our final factor solution is derived from an EFA of the entire sample (*N*=2777) and is presented in Table 6. Each step in the process of deriving it is described in detail in the following three subsections.

Model Development.
To develop the initial model of the Spring 2015 data, we computed factor solutions using both Principal Component Analysis (PCA) and Principal Factors Analysis ("maximum likelihood") in SPSS. We had reason to expect that any underlying affective factors are likely to be correlated with one another, so we opted to compute rotated factor solutions via the oblique rotation method "direct oblimin." Rotated factor solution methods allow inter-factor correlation rather than forcing orthogonality on the solution, and thus allow for extraction of correlated factors.

In order to simplify interpretation, we reverse-coded a number of affective questions (marked with asterisks in Table 1) so that they were consistently ordered from negative to positive. All standard metrics for sampling adequacy[3] and multicolinearity[4] were well within acceptable ranges for this dataset, indicating suitability for factor analysis. An initial factor solution showed the effort variable to be highly crossloading (simultaneously loading onto multiple factors), which limited our ability to obtain a clean solution, and so this variable was treated as a standalone factor and removed from the EFA at the outset.

The scree plot for the remaining 25 variables shows a single strong inflection point at four factors, suggesting that extraction of three factors is justified. We choose to extract only factor loadings with absolute values greater than 0.32, corresponding to retention of variables whose variance overlaps by ≥10% with other variables loading onto that factor (Costello and Osbourne 2005).

Six of the 25 input variables do not load at >0.32 onto any of the three factors, and an additional two ("Math is Fun (1) to Scary (4)" and the Likert-scale statement "Numerical skills are important to the understanding of science") crossloaded onto factors 1 and 2 and tended to vary in which factor they loaded

---

[2] For an accessible introduction to Factor Analysis and choices of methods and parameters, including all that are reported/chosen here, see *Discovering Statistics Using IBM SPSS*, Chapter 17 (Field 2013)

[3] The Kaiser-Meyer-Olkin measure of sampling adequacy was always >0.8 and Bartlett's Test of Sphericity was highly significant (*p*<0.001).

[4] The determinant of the correlation matrix was 0.002 at minimum, well above nominal cutoff of 0.00001), and all of the off-diagonal elements in the correlation matrix are <0.8

onto across analysis methods. All seven of these variables were removed before deriving a final EFA solution for the Spring 2015 data, which contains three clean (free of crossloading) factors, each with 5 to 6 individual variables loading onto them.

Model Validation.

To test the structure of the initial model, we conducted a Confirmatory Factor Analysis (CFA) on the remaining data (Fall 2015+Spring 2016, N=1289) using the EFA factor solution from the Spring 2015 subsample (N=1488). The overall fit of the CFA was found to be acceptable, validating the use of composite EFA variables in our final regression models (described in the next section). The global fit indices for the CFA model with the three factors from the Spring 2015 EFA fit are reported in Table 5 with recommendations for acceptable and good fit values for these parameters in parentheses (Hu and Bentler 1999).

Table 5
Confirmatory Factor Analysis Model Metrics

Metric	Value for QuaRCS F15+S16 dataset	Recommendations for acceptable values
Standardized Root Mean Square Residual (SRMR)	0.045	<0.08 = good fit
Root Mean Square Error of Approximation (RMSEA)	0.044	<0.05 = acceptable
Comparative Fit Index	0.971	>0.95 = good
Tucker Lewis Index	0.966	>0.95 = good

Notes: Summary of the Confirmatory Factor Analysis (CFA) model, with descriptions of generally-accepted ranges for high quality fits.

As a final test of the validity of the factor structure, EFA solutions for each of the two subsamples were compared to one another and to the solution for the entire sample. The same three factors are consistently present across these analyses, with the same variables loading onto each one. The maximum variation in EFA factor loadings across the three samples is 0.14 for Maximum Likelihood analysis and 0.12 for PCA. We chose to extract our final solution using the Maximum Likelihood method because the calculation attempts to maximize overlapping variance rather than total variance, thus allowing for some error variance and variance unique to each factor. Although this is inarguably the more correct approach, we note that the PCA solution is virtually identical to the Maximum Likelihood solution.

Final Affective Factor Solution.

Given the stability of the solution to the tests described in the previous section, we elect to compute the final composite variable loadings (shown in Table 6) from an EFA of the entire sample (N=2777), as it has the most statistical power. One variable, the Likert-scale statement "I am satisfied with my current level of numerical ability," appears only in the Fall 2015 and Spring 2016 data and was

not included in the final EFA (though its loadings in the Fall 2015+Spring 2016 subsample are shown in red in Table 6).

Table 6
Factor Loadings

No.	Variable	Factor 1	2	3
7	I feel confident using numbers in my non-math courses	**0.89**	-0.08	0.00
8	I feel confident using numbers in my everyday life	**0.83**	-0.05	0.02
9	I am satisfied with my current level of numerical ability[§]*	**0.64**	-0.12	-0.01
6c	Math is: Easy (1) to Hard (4)	**0.60**	0.06	-0.02
1	Overall, how difficult were the questions in this survey?	**0.60**	0.04	-0.01
Calc.	Average confidence ranking	**0.55**	0.11	0.09
4	How frequently do you do calculations in your everyday life?	-0.06	**0.72**	0.02
5	How frequently do you encounter graphs and tables in your daily life?	-0.08	**0.65**	-0.02
3	How frequently do you encounter situations similar to the questions on this survey?	0.00	**0.58**	0.02
6b	Math is: Useful (1) to Useless (4)	0.12	**0.51**	0.03
10	Numerical skills are important to my daily life	0.27	**0.44**	0.00
6a	Math is: Interesting (1) to Boring (4)	0.24	**0.41**	-0.03
18b	I chose my major because: I feel that it will help me get a job I will enjoy after graduation	-0.05	0.04	**0.60**
18c	I chose my major because: I am good at it	0.07	-0.05	**0.53**
18d	I chose my major because: I feel that it will help me get a well-paying job after graduation	-0.01	0.07	**0.49**
18a	I chose my major because: I like the subject	0.01	-0.02	**0.44**
18g	I chose my major because: I'm not sure yet	0.02	0.03	**-0.43**
	*Crossloading Items**			
6d	*Math is: Fun (1) to Scary (4)*	*0.43*	*0.33*	*-0.06*
9	*Numerical skills are important to the understanding of science*	*0.29*	*0.32*	*0.10*
	*Variables That Do Not Load Strongly Onto These Factors**			
18e	*I chose a major that would avoid math as much as possible*	*-0.23*	*-0.12*	*0.31*
18f	*I chose a major that would avoid writing as much as possible*	*0.03*	*0.03*	*0.18*
15d	*I chose this course because: it sounded easy*	*0.07*	*0.02*	*0.12*
15c	*I chose this course because: it sounded interesting*	*0.00*	*0.19*	*0.18*
15g	*I chose this course because: Other*	*0.02*	*0.10*	*0.01*
3	*Which of the following best describes your calculator usage while taking this survey?*	*0.03*	*-0.09*	*-0.02*

Notes: Summary of the final Exploratory Factor Analysis (EFA) model. Factor loadings above 0.32 are shown in bold, and those below are shown in grey for completeness. Variables not included in the final model are given in italics at the bottom of the table, and were dropped for reasons described in detail in the text.

[§]Additionally, an error in question 9 made data for this variable unusable in the Spring 2015 data. Its loading for the two semesters in which it was included is shown here in red, but the variable was dropped before computing the final factor scores used in the regression model.

*If included in the solution (altering the numbers in the rest of the table slightly)

The loadings in Table 6 represent the Pattern Matrix for the final factor solution. A Pattern Matrix gives the unique loadings for each variable onto each factor or, more accurately, the regression coefficients for each variable onto each factor. The three extracted factors are composed of between 5 and 6 variables each. For completeness, loadings for the affective variables that were removed in the process of computing the initial solution are given at the bottom of Table 6. This analysis argues strongly for collapsing the eighteen affective variables listed at the

top of Table 6 into three composite variables, which we name and describe in the Analysis and Interpretation section.

Table 7
Factor Correlation Matrix

Factor	1	2	3
1	1.0	0.54	0.20
2	0.54	1.00	0.12
3	0.20	0.12	1.00

Notes: Pearson's correlation coefficients between identified factors. The high interfactor correlations, particularly between factors 1 and 2, justify our choice to compute a rotated factor solution.

The factor correlation matrix quantifies the overall relationship between factors and is given in Table 7. Factors 1 and 2 are correlated at the 0.54 level, suggesting that our choice of orthogonal rotation was justified. The meaning of these factors is explored in the next section.

Analysis and Interpretation

Dimension Reduction

Definition of Composite Attitudinal Variables.
As the purpose of our factor analyses was to reduce measured variables into a composite set of more profound underlying variables, we aim to understand their composition and meaning, and to assign descriptive names to each one.

Factor 1 is composed of variables that measure perceived difficulty of the assessment, students' average confidence in in their answers, their confidence in using numbers in their daily life and coursework, and their feelings about where "doing mathematics" lies on a 4 point scale between Easy and Hard. The fact that these variables load onto a common factor is unsurprising, as all are related to a student's underlying sense of self-efficacy, specifically in regard to mathematics. Therefore, we have chosen to call this composite factor "numerical self-efficacy".

The variables loading onto Factor 2 address students' perceptions of the utility of mathematics and the frequency with which they encounter calculations, graphs, and situations similar to the problems in the QuaRCS survey in their daily lives. We have chosen to call this factor "numerical relevancy"

Of the three factors revealed in this analysis, Factor 3 is the most difficult to interpret. It is composed of five of the eight possible choices to the question "I chose my major because… (choose all that apply)". The simplest relationship between these variables appears to be that they all indicate a level of academic maturity, so we call this factor "academic maturity".

Table 8 provides details about the correlations between the three named factors and QuaRCS score. As the variables are not normally distributed (Fig. 6), we report Spearman's Rho correlations in lieu of Pearson's correlation

coefficients, though we note that these two values differ by less than 0.03 in each case.

Table 8
Final Affective Factors

Factor	Factor description	Spearman's rho correlation with QuaRCS Score	Sig.
Factor 1	Numerical Self-Efficacy	0.538	<0.001
Factor 2	Numerical Relevancy	0.209	<0.001
Factor 3	Academic Maturity	0.117	<0.001

Notes: This table lists the names that we have assigned to each of the factors identified by the EFA analysis, as well as their correlations with QuaRCS score and the significance of these correlations.

Individual student scores for each of these composite variables are calculated using the Anderson-Rubin method, which attempts to extract factor scores that are uncorrelated with one another (DiStefano, Zhu and Mîndrilă 2009). Histograms of student scores for each of the three composite variables are given in Figure 6. The self-efficacy variable is markedly double peaked and asymmetric, suggesting perhaps a small population of students with high math self-efficacy and a larger population of students with low math self-efficacy. The numerical relevancy variable, on the other hand, is broad and single peaked, suggesting a wide range of beliefs. The academic maturity variable does not follow any clear underlying distribution.
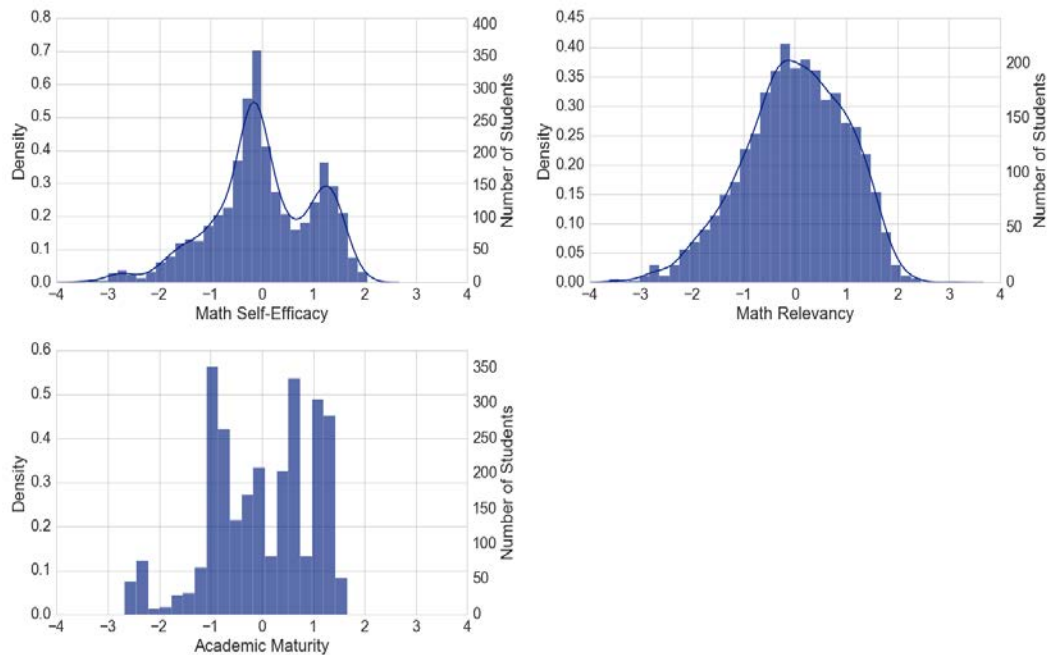

Figure 6. Distributions of individual student scores on the three composite variables calculated from the EFA. The *x*-axis values represent standard deviations from the mean of the distribution for each factor score.

Regression Model

Regression Model Development.

All regression modeling was done using the built-in linear regression function in SPSS. Variables were entered hierarchically into the model in tiers according to question classification (as composite, affective, academic demographic, or basic demographic). Since order matters in a hierarchical model, these tiers were ordered according to the average effect size of their variables. The first tier in the initial model consisted of the three composite affective variables developed in the Results section, and the second tier was composed of all affective variables that were orphaned in that analysis, either due to crossloading or not appearing in the final factor solution. The third and fourth tiers consisted of all academic and basic demographic variables that were shown to correlate with QuaRCS score.

The initial factor solution revealed that many of the variables shown in the Study Design and Methods section to correlate significantly with QuaRCS score add no significant explanatory power to the regression model. The most likely explanation for this is that the correlation of these variables with QuaRCS score is driven by their correlation with other, more significant, variables. These variables were removed before computing the final regression model, described in the following subsection. Disaggregation of the variables driving the correlations between QuaRCS score and demographic variables is left for future work.

Final Regression Model.

The output of the final regression model, which consists of four tiers, is described in Table 9. The percent of variance in QuaRCS score (R^2) accounted for as each tier is added to the model is given in the rightmost column, and reveals that the four-tiered model can explain 49% of the variance in score. However, the vast majority of the explanatory power in this model is held in the first two affective variable tiers, which together account for 48.3% of the variance.

More specifically, the three composite variables alone account for 32.4% of total variance. Of the affective variables not included in the first tier composites, only the effort and calculator usage variables add additional explanatory power to the model, and together they account for an additional 15.9%. Although the interrelation between individual correlation coefficients and the final R^2 value is complicated by correlations between variables within and among tiers, the high raw correlation between score and effort level ($R=0.44$) suggests that it is responsible for most of the additional explained variance in the second tier of the model. Indeed, a model with the three composite variables in the first tier and effort alone in the second tier results in an R^2 value of 47.2%, suggesting that the calculator usage variable contributes just over 1% to the final model. Four academic demographic variables add only an additional 0.6%, and two basic

demographic variables just 0.1% beyond that. Scores predicted by the regression model are shown plotted against actual QuaRCS scores in Figure 7.

Table 9
Final Regression Model

Factor/variable	Standardized beta coefficient	Significance	Pearson's correlation coefficient (partial)	Model R^2
Model 1: Composite Affective Variables from Dimension Reduction				
Numerical Self-Efficacy	0.42	<0.001	0.43	
Numerical Relevancy	0.15	<0.001	0.16	0.324
Academic Maturity	0.07	<0.001	0.09	
Model 2: Model 1+ Affective variables not appearing in or removed from factor solution				
Effort	0.39	<0.001	0.44	0.483
Calculator Usage	-0.07	<0.001	-0.10	
Model 3: Model 2+ Academic Variables				
Chose my major because: Other	0.05	<0.001	0.07	
Numerical skills are important to my daily life	0.09	<0.001	0.09	0.489
Humanities major	0.05	<0.001	0.07	
STEM major	0.06	<0.001	0.09	
Model 4: Model 3+ Basic Demographics				
Learning disability	-0.03	0.02	-0.05	0.493
Age	0.03	0.02	0.05	

Notes: Summary of the regression model that produces Figure 7. Variables were entered in four tiers, and the additional descriptive power added by the variables in each tier can be seen in the change in the R^2 value as that tier is added to the model, where R^2 represents the proportion of variation in QuaRCS score explained by the model.
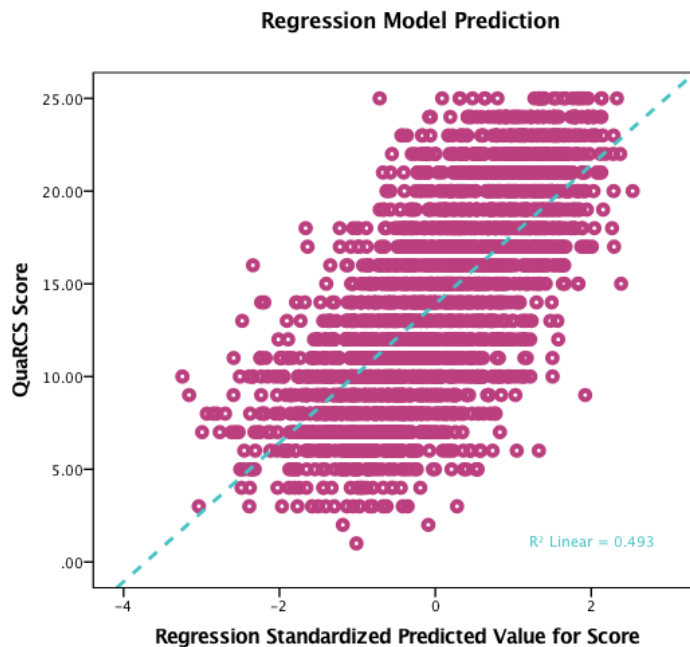

Regression Model Prediction

Figure 7. Standardized predicted score based on a regression model of affective and demographic variables vs. actual score on the QuaRCS. The supplemental QuaRCS questions explain nearly 50% of the observed variance in score.

Instrument Revisions

The analyses described in this paper allowed us to reexamine the purpose and utility of each non-quantitative question on the assessment. A number of revisions, additions and removals were clearly warranted, and we describe them in this section.

Removal of Non-Informative and Redundant Questions.
Questions regarding intended college coursework in science and mathematics proved to be most indicative of whether or not a student was a STEM major, which we can and do probe more directly by asking for students' majors. These questions have been removed from the assessment. The question asking students to specify their most recent mathematics course was shown to have no relationship to assessment score and has also been removed.

The EFA solution consistently showed between 5 and 6 variables loading onto each factor, more than is necessary for a stable solution. As we wished to trim the assessment down to only essential questions in order to limit fatigue, we removed several that we deemed less informative. These were: perceived difficulty of the assessment, the Likert Statement "Numerical skills are important to my everyday life," and the question asking students to rate doing math on a scale between interesting and boring. The difficulty question is the lowest loading non-calculated variable onto the Numerical Self-Efficacy factor, and the others are the lowest two loading variables onto the Numerical Relevancy factor.

Two crossloading variables were also removed, namely the question asking students to rank doing math on a scale between fun and scary and the Likert statement "Numerical skills are important to the understanding of science."

Question Revisions.
We also simplified, clarified or revised the wording and/or answer choices of several questions. First, we elected to drastically simplify the list of majors to include only those that are most important or interesting in the context of quantitative reasoning and those that were shown to correlate with score on the assessment. We collapsed Arts and Humanities into a single category, removed the "Trade-specific" and "General studies" categories, and removed the long list of example majors for each category. Although we expect an increase in the number of students specifying "other" for this question as a result, we also believe that this simplification will decrease cognitive load and student frustration at being confronted with a very long and complicated list of choices. We also split the "Science" category into "Biological and Health Sciences" and "Physical Sciences," which will allow us to explore differences between these groups in the future.

Table 10
Question Modifications

Original Question	Revised Question(s)
Where would you put mathematics (including: doing calculations, reading graphs and tables, reasoning with numbers, etc.) on the following scales between two opposite adjectives *Interesting to Boring* *Easy to Hard* *Useful to Useless* *Fun to Scary*	Think of several instances in which you have had to use math (including: doing calculations, reading graphs and tables, reasoning with numbers, etc.) to solve a problem ***in your daily life/for school [separate questions].*** How do these experiences usually rate on the following scales between two opposite adjectives? *Easy to Hard* *Fun to Boring* *Useful to Pointless* *Straightforward to Confusing* *Not stressful to Stressful*
None (new question)	Please indicate the level of your anxiety in the following situations. • when a teacher uses an equation to explain something • when I have to use a <u>graph</u> to answer a homework problem • when I have to calculate a tip • when someone uses numbers or statistics to make an argument that I disagree with *Not at all* *a little* *a fair amount* *a lot*
Please select your major or majors from the list below. *Humanities (e.g. English, Literature, Philosophy, Religion, Foreign Language)* *Arts (e.g. Art, Music, Dance, Film, Theater, Creative Writing)* *Social Sciences (e.g. Economics, History, Political Science, Psychology, Sociology, Anthropology, International Relations, Geography, Linguistics, etc.)* *Education* *Science (e.g. Physics, Chemistry, Biology, Geology and sub-disciplines)* *Engineering, Mathematics or Computer Science* *Business-related (e.g. Business, Marketing, Management)* *Health-related (Nursing, Pharmacy, Nutritional Science, Public Health, Exercise Science, etc.)* *Trade-specific (e.g. Architecture, Agriculture, Law, Justice, Library Science, Retail, Family and Consumer Sciences, Construction Trades, etc.)* *Journalism* *General Sudies* *Undecided* *Other*	Please select your major or majors from the list below. *Arts and Humanities* *Social and Behavioral Sciences* *Biological and Health Sciences* *Physical Sciences* *Engineering, Mathematics or Computer Science* *Education* *Business-Related* *Journalism* *Other* *Undecided*

Notes: This table summarizes the changes that we have made to the QuaRCS based on results from this analysis with the original question wording (where relevant) at left and the new wording at right. Answer choices are given in italics.

Our hypothesis about the crossloading nature of the opposite adjective scale fun to scary is that these adjectives are not always perceived as opposites. We revised this scale to be fun to boring, which we believe are more commonly

perceived as opposites. We also changed the adjective "useless" to "pointless" to ensure that the scale is read carefully.

Question Additions.
We added several questions to the assessment in an attempt to probe math anxiety, which is not effectively measured with the original assessment. These include two new opposite adjective scales: Straightforward to Confusing and Not Stressful to Stressful. We also added an entirely new question about students' level of anxiety in various situations.

In attempting to probe math anxiety, the distinction between real world and academic/classroom contexts becomes increasingly important, and so we have also split the opposite adjective question into two: one that asks students to rate using math "in [their] daily life" and one "for school." The new situational anxiety question also includes two academic and two daily life situations. A summary of the added and revised question is provided Table 10.

Conclusions

In this paper, we've explored correlations between students' scores on the 25 quantitative QuaRCS questions and various demographic, academic, and attitudinal variables measured by the non-quantitative questions on the assessment. Analyzing the distributions of student scores on the assessment according to their answers to the non-quantitative questions reveals a number of groups that deviate significantly from the overall population and/or from one another.

In total, 45 separate variables show statistically significant effects on QuaRCS score. However, we expect neither that these 45 variables measure independent quantities nor that they have equal explanatory power in predicting QuaRCS score. Indeed, we show that seventeen of the raw variables can be collapsed into three much more profound underlying factors, which we have identified as "numerical self-efficacy," "numerical relevancy," and "academic maturity." These three variables alone explain 32% of the variance in QuaRCS score.

Of the 28 remaining variables with significant effects on QuaRCS score, only eight of them add explanatory power to a regression model predicting that score: two additional affective variables, four academic demographic variables, and two basic demographic variables. Of these, the additional affective variables account for most of the added explanatory power in the final model, increasing explained variance from 32% to 48%. The academic and basic demographic variables add less than 1% to the model.

The lack of appearance of race, gender, and status as an education major in the final regression solution suggests that the well-documented differences in numerical assessment scores associated with these variables may be driven by underlying differences in attitudes toward mathematics. The most significant conclusion of this work is thus that correlations between basic demographic variables and quantitative ability measures are significant but not strongly predictive. Attitudinal measures, on the other hand, are *very* strongly predictive. Therefore, it is possible that differences in score associated with at least some demographic variables are driven by underlying differences in affect.

We believe that these results point to remediation of attitudes regarding mathematics as an important step on the road to a quantitatively literate population of college graduates. The implication of this result for classroom practitioners is that working to improve learners' attitudes and self-efficacy may be every bit as important as remediating numerical skills if we wish to improve our students' quantitative reasoning abilities, if not more so.

References

Betz, Nancy E. 1978. "Prevalence, Distribution, and Correlates of Math Anxiety in College Students." *Journal of Counseling Psychology* 25 (5): 441–448. https://doi.org/10.1037/0022-0167.25.5.441.

Betz, Nancy E., and Gail Hackett. 1983. "The Relationship of Mathematics Self-Efficacy Expectations to the Selection of Science-Based College Majors." *Journal of Vocational Behavior* 23 (3): 329–345. https://doi.org/10.1016/0001-8791(83)90046-5.

Bong, Mimi and Einar M. Skaalvik. 2003. "Academic Self-Concept and Self-Efficacy: How Different are they Really?" *Educational Psychology Review* 15 (1): 1–40. https://doi.org/10.1023/A:1021302408382.

Bridgeman, Brent and Cathy Wendler. 1991. "Gender Differences in Predictors of College Mathematics Performances and in College Mathematics Course Grades." *Journal of Educational Psychology* 83 (2): 275–284. https://doi.org/10.1037/0022-0663.83.2.275.

Chan, Lorna K.S. and Kerry Dally. 2001. "Learning Disabilities and Literacy and Numeracy Development." *Australian Journal of Learning Difficulties* 6 (1): 12–19. https://doi.org/10.1080/19404150109546652.

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106: 1691–1730. https://doi.org/10.1086/321299.

Costello, Anna B. and Jason Osborne. 2005. "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis." *Practical Assessment Research and Evaluation* 10 (7).

DeBoer, George E. 2000. "Scientific Literacy: Another Look at Its Historical and Contemporary Meanings and Its Relationship to Science Education Reform." *Journal of Research in Science Teaching* 37 (6): 582–601. https://doi.org/10.1002/1098-2736(200008)37:6<582::AID-TEA5>3.0.CO;2-L.

DiStefano, Christine, Min Zhu, and Diana Mîndrilă. "Understanding and Using Factor Scores: Considerations for the Applied Researcher." *Practical Assessment, Research and Evaluation* 14 (20).

Durrani, Naureen, and Vicki Tariq. 2009. "Relationships between Undergraduates' Mathematics Anxiety and Their Attitudes towards Developing Numeracy Skills and Perceptions of Numerical Competence," 787–794. http://library.iated.org/view/DURRANI2009REL.

Eccles, Jacquelynne S. 1994. "Understanding Women's Educational and Occupational Choices: Applying the Eccles et al. Model of Achievement-Related Choices." *Psychology of Women Quarterly* 18 (4): 585–609. https://doi.org/10.1111/j.1471-6402.1994.tb01049.x.

Eagan, Kevin, Ellen Bara Stolzenberg, Abigail K. Bates, Melissa C. Aragon, Maria Ramirez Suchard, and Cecilia Rios-Aguilar. 2015. *The American Freshman: National Norms Fall 2015.* Los Angeles: Higher Education Research Institute, UCLA.

Field, Andy. 2013. *Discovering Statistics Using IBM SPSS Statistics, 4th Edition.* London-Thousand Oaks-New Delhi-Singapore: Sage Publications.

Follette, Katherine B., Donald W. McCarthy, Erin Dokter, Sanlyn Buxner, and Edward Prather. 2015. "The Quantitative Reasoning for College Science (QuaRCS) Assessment, 1: Development and Validation," *Numeracy* 8 (2): Article 2. https://doi.org/10.5038/1936-4660.8.2.2.

Gilbert, Lisa A., Jennifer Stempien, David A. McConnell, David A. Budd, Katrien J. van Der Hoeven Kraft, Ann Bykerk-Kauffman, Megan H. Jones, Catharine C. Knight, Ronald K. Matheney, Dexter Perkins, and Karl R. Wirth. 2012. "Not Just 'Rocks for Jocks': Who Are Introductory Geology Students and Why Are They Here?" *Journal of Geoscience Education* 60 (4): 360–371. https://doi.org/10.5408/12-287.1.

Hu, Li-tze and Peter M. Bentler. 1999. "Cutoff Criteria for Fit Indices in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling: A Multidisciplinary Journal* 6 (1): 1-55. https://doi.org/10.1080/10705519909540118.

Impey, Chris, Sanlyn Buxner, Jessie Antonellis, Elizabeth Johnson, and Courtney King. 2011. "A Twenty-Year Survey of Science Literacy Among College Undergraduates." *Journal of College Science Teaching* 40 (4): 31–37.

Johnson, Laurence F. 1996. "Developmental Performance as a Predictor of Academic Success in Entry-Level College Mathematics." *Community College Journal of Research and Practice* 20 (4): 333–344. https://doi.org/10.1080/1066892960200404.

Kelly, Ronald R., and Martha G. Gaustad. 2006. "Deaf College Students' Mathematical Skills Relative to Morphological Knowledge, Reading Level, and Language Proficiency." *Journal of Deaf Studies and Deaf Education* 12 (1): 25–37. https://doi.org/10.1093/deafed/enl012.

Ma, Liping. 1999. *Knowing and Teaching Elementary Mathematics: Teachers' Understanding of Fundamental Mathematics in China and the United States.* Mahwah, N. J.: Lawrence Erlbaum Associates.

Manduca, Cathryn A., Eric Baer, Greg Hancock, R. Heather Macdonald, Sam Patterson, Mary Savina, and Jennifer Wenner. 2008. "Making Undergraduate Geoscience Quantitative." *Eos* 89 (16): 149–150. https://doi.org/10.1029/2008EO160001.

Matthews, Kelly E., Peter Adams, Merrilyn Goos, and Kelly Elizabeth Matthews. 2009. "Putting It into Perspective: Mathematics in the Undergraduate Science Curriculum." *International Journal of Mathematics Education in Science and Technology* 40 (7): 891–902. https://doi.org/10.1080/00207390903199244.

Matthews, Michael E., and Walter I. Seaman. 2007. "The Effects of Different Undergraduate Mathematics Courses on the Content Knowledge and Attitude towards Mathematics of Preservice Elementary Teachers." *IUMPST: The Journal* 1 (July): 64–72.

McGlaughlin, Sean M., Andrew J. Knoop, and Gregory A. Holliday. 2005. "Differentiating Students with Mathematics Difficulty in College: Mathematics Disabilities vs. No Diagnosis." *Learning Disability Quarterly* 28: 223–232. https://doi.org/10.2307/1593660.

Moakler, Martin W., and Mikyong Minsun Kim. 2014. "College Major Choice in STEM: Revisiting Confidence and Demographic Factors." *Career Development Quarterly* 62 (2): 128–142. https://doi.org/10.1002/j.2161-0045.2014.00075.x.

Ogbu, John U. 1990. "Minority Status and Literacy in Comparative Perspective." *Daedalus* 119 (2): 141–168.

Pierce, Rebecca L. 2006. "SATS at Ball State University: Approaches and Attitudes." In A. Rossman and B. Chance (eds.), *Working Cooperatively in Statistics Education: Proceedings of the Seventh International Conference on Teaching Statistics,* Salvador Brazil. [CDROM}. Voorburg, The Netherlands: International Statistical Institute.

Raudenbush, Stephen and Rafa Kasim. 1998. "Cognitive Skill and Economic Inequality: Findings from the National Adult Literacy Survey." *Harvard Educational Review* 68 (1): 33–80. https://doi.org/10.17763/haer.68.1.1j47150021346123.

Rousselle, Laurence and Marie-Pascale Noël. 2005. "Basic Numerical Skills in Children with Mathematics Learning Disabilities: A Comparison of Symbolic vs. Non-symbolic Number Magnitude Processing." *Cognition* 102: 361–395. https://doi.org/10.1016/j.cognition.2006.01.005.

Sax, Linda J. 1994. "Mathematical Self-Concept: How College Reinforces the Gender Gap. *Research in Higher Education* 35 (2): 141–166. https://doi.org/10.1007/BF02496699.

Sax, Linda J., M. Allison Kanny, Tiffani A. Riggers-Pielh, Hannah Whang, and Laura N. Paulson. 2015. "'But I'm Not Good at Math': The Changing Salience of Mathematical Self-Concept in Shaping Women's and Men's STEM Aspirations." *Research in Higher Education* 56 (8): 813–842. https://doi.org/10.1007/s11162-015-9375-x.

Schau, Candace, Joseph Stevens, Thomas L. Dauphinee, and Ann Del Vecchio. 1995. "The Development and Validation of the Survey of Attitudes Toward Statistics." *Educational & Psychological Measurement*, 55 (5): 868–876. https://doi.org/10.1177/0013164495055005022.

Sparks, Richard L., and Benjamin J. Lovett. 2009. "College Students With Learning Disability Diagnoses: Who Are They and How Do They Perform?" *Journal of Learning Disabilities* 42 (6): 494–510. https://doi.org/10.1177/0022219409338746.

Spencer, Steven J., Claude M Steele, and Diane M. Quinn. 1999. "Stereotype Threat and Women's Math Performance." *Journal of Experimental Social Psychology* 35 (1): 4–28. https://doi.org/10.1006/jesp.1998.1373.

Steele, Claude M., and Joshua Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69 (5): 797–811. https://doi.org/10.1037/0022-3514.69.5.797.

Taub, Gordon E., Nicholas Benson, and Judit Szente. 2014. "Improving Mathematics: An Examination of the Effects of Specific Cognitive Abilities on College-Age Students' Mathematics Achievement." *International Journal for the Scholarship of Teaching and Learning* 8 (2): Article 8. https://doi.org/10.20429/ijsotl.2014.080208.

Voyer, Daniel and Susan D. Voyer. 2014. "Gender Differences in Scholastic Achievement: A Meta-Analysis." *Psychological Bulletin* 140 (4): 1174–1204. https://doi.org/10.1037/a0036620.

Wang, Ming-Te. 2012. "Educational and Career Interests in Math: A Longitudinal Examination of the Links between Classroom Environment, Motivational Beliefs, and Interests." *Developmental Psychology* 48 (6): 1643–1657. https://doi.org/10.1037/a0027247.

Wang, Xueli. 2013. "Why Students Choose STEM Majors: Motivation, High School Learning, and Postsecondary Context of Support." *American Educational Research Journal* 50 (5): 1081–1121. https://doi.org/10.3102/0002831213488622.

Xie, Yu and Alexandra A. Killewald. 2012. *Is American Science in Decline?* Cambridge, Massachusetts: Harvard University Press. https://doi.org/10.4159/harvard.9780674065048.