

2018

# Using the Quantitative Literacy and Reasoning Assessment (QLRA) for Early Detection of Students in Need of Academic Support in Introductory Courses in a Quantitative Discipline: A Case Study

Nathan D. Grawe  
Carleton College, [ngrawe@carleton.edu](mailto:ngrawe@carleton.edu)  
Kristin O'Connell  
Carleton College, [koconnell@carleton.edu](mailto:koconnell@carleton.edu)

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

 Part of the [Curriculum and Instruction Commons](#), [Economics Commons](#), and the [Educational Assessment, Evaluation, and Research Commons](#)

## Recommended Citation

Grawe, Nathan D., and Kristin O'Connell. "Using the Quantitative Literacy and Reasoning Assessment (QLRA) for Early Detection of Students in Need of Academic Support in Introductory Courses in a Quantitative Discipline: A Case Study." *Numeracy* 11, Iss. 1 (2018): Article 5. DOI: <https://doi.org/10.5038/1936-4660.11.1.5>

---

# Using the Quantitative Literacy and Reasoning Assessment (QLRA) for Early Detection of Students in Need of Academic Support in Introductory Courses in a Quantitative Discipline: A Case Study

## Abstract

As the number of young people attending college has increased, the diversity of college students' educational backgrounds has also risen. Some students enter introductory courses with math anxiety or gaps in their quantitative training that impede their ability to master or even grasp relevant disciplinary content. Too often professors learn of these anxieties and gaps only during the post mortem of the first midterm. By that time, a good portion of a student's grade is determined and successful recovery may be impossible. During the 2016-17 academic year, the Department of Economics at Carleton College ran a pilot project using the Quantitative Literacy and Reasoning Assessment (QLRA) as a pre-course diagnostic tool. Results show that the QLRA predicts student grades even after controlling for other SAT/ACT math scores and overall GPA. This finding suggests that quantitative reasoning is an important input into success in Principles of Economics (both Macro and Micro). When the QLRA alone is used to predict success in a course (as defined by either a grade of C- or better, or a grade of B- or better), we find that we could nearly always pick out students who were on the way to sub-par performance. On the other hand, the tool has a fairly high false positive rate; almost half of students identified as "at risk" based on QLRA performance went on to earn a successful grade in the course. In total, we argue that the QLRA may be a useful and inexpensive early-warning device for introductory courses in economics; it may be worth exploring a similar use of the instrument in other disciplinary settings where introductory courses require quantitative reasoning.

## Keywords

economics, assessment, Quantitative Literacy and Reasoning Assessment

## 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

Nathan D. Grawe is Professor of Economics and former director of the Quantitative Inquiry, Reasoning, and Knowledge (QuIRK) initiative at Carleton College.

Kristin O'Connell is Evaluation/Education Specialist for the Science Education Resource Center and Quantitative Skills Coordination for TRiO Student Support Services at Carleton College.

## Introduction

In its LEAP campaign, the Association of American Colleges & Universities (2007) lists quantitative literacy/reasoning among its “Essential Learning Outcomes” which support students’ success in the 21<sup>st</sup> century. As Vacher (2011) notes, the LEAP conception of quantitative reasoning can take many forms as colleges seek to prepare a numerate generation. At our institution, we see the commitment to quantitative reasoning playing out in at least two ways. First, students will encounter and have need for quantitative arguments in personal and public life and should be empowered to engage those issues. Second, quantitative fluency is essential for many potential career paths, some of which begin within college majors. We hope that all of our students have the quantitative foundation required to complete whatever major is appropriate for them. However, it stands to reason that any discipline involving quantitative modeling and argument may appear closed to students with weaker quantitative reasoning skills despite being interested and otherwise qualified.

With such questions in mind, some members of the Economics Department at Carleton College have been studying the success of students in our introductory course sequence. We suspect that experiences in these gateway courses are the primary impediments to many students’ pursuit of the major. Of course, the vast majority of students do just fine in these courses and so can freely contemplate continuing on into the major. For example, in 2016-17 only 3 percent of our students earned a grade below C-, only 18 percent fell short of B- work, and only 30 percent earned less than a B. With so many students achieving strong grades, we do not struggle to generate sizeable numbers of majors. Still, when one-fifth of students arrive at grades that might discourage them from even contemplating a major in the discipline, we wonder if there are ways we might address barriers that lead to unsatisfactory outcomes.

We hypothesize that many students who do poorly in introductory economics courses struggle because they lack foundational quantitative skills. While other students are learning economic models, these students are simultaneously backfilling mathematics while attempting to acquire economics. Often we identify at-risk students at the time of the first midterm. They perform poorly and, through examination of the paper along with conversations with the students, we suspect that weak quantitative foundations explain much of the struggle. At that point, with one-third or more of the term already gone and a good portion of the course grade already locked in at a low level, it seems unlikely that any intervention could succeed. Indeed, any hope of success in mitigating weak quantitative foundations developed over more than a decade in the K-12 system relies on early detection.

In this study we examine whether the free-to-use Quantitative Literacy and Reasoning Assessment (QLRA) developed by Gaze et al. (2014) can be used to identify students at risk for low performance in introductory economics courses.<sup>1</sup> We asked all students taking either Principles of Macroeconomics or Principles of Microeconomics during the 2016-17 academic year to complete the QLRA, and we then tracked subsequent course performance to see how well the instrument might serve as an early warning for poor final grades. We found that the QLRA does predict course performance. Moreover, it has predictive power even after controlling for math ACT and SAT scores and cumulative GPA. Apparently, quantitative reasoning as measured by the QLRA is meaningfully connected with performance in our introductory economics curriculum. We conclude by discussing how we plan to use this information to design interventions that mitigate quantitative reasoning skills gaps. While our context differs from that at many institutions, we hope many will find our process of using the QLRA to be useful as a model.

## Quantitative Literacy and Inclusion

Quantitative skills play a uniquely strong role in determining the insiders and outsiders of higher education. Bailey, Jeong, and Cho (2010) and Hacker (2012) note how performance in formal mathematics limits high school graduation, college matriculation, and college completion. Moreover, Wang, Degol, and Ye (2015) note that math achievement also serves as a filter that limits entry by underrepresented groups like women into STEM fields. While it may be tempting to view these findings as evidence of the importance of simple math and/or computational literacy, Wang et al. find that math task value—the propensity to enjoy math and see it as “useful in everyday problems”—is of at least as much importance as raw math aptitude. Of course, to readers of this journal, this important math task value has significant overlap with broader notions of quantitative reasoning.

These quantitative barriers are of growing importance as higher education expands its reach into underserved communities. According to the National Center for Education Statistics (NCES), over the last four decades we have witnessed the share of high school completers attending college increase from about half to nearly three-quarters (NCES 2016, Table 302.1). While the share of the increase attributable to growth in enrollments at two-year schools is greater than their enrollment share, matriculation rates to four-year institutions are up more than 25 percent. This growth in attendance at institutions of higher education reflects expansions into previously less-recruited communities.

---

<sup>1</sup> The instrument is available by emailing Eric Gaze at [egaze@bowdoin.edu](mailto:egaze@bowdoin.edu).

In these communities, students often face an opportunity gap with respect to rigorous K-12 preparation. This gap disproportionately affects low-income (LI) and first-generation-in-college (FG) students (see Warburton et al. 2001; Kuh et al. 2006; Engstrom and Tinto 2008.) In our sample (described below), LIFG students at Carleton had lower course grades, SAT math, and QLRA scores. Early detection of gaps in foundational quantitative reasoning skills coupled with early intervention is thus a matter of inclusion; lack of opportunity prior to college should not close the door to a major. Many disciplines and funding agencies have called for this type of reflective institutional accountability (Asai and Bauerle 2016) to ensure that all students accepted to our institutions have pathways to succeed.

## Institutional Context

While we believe our experience can provide a useful example for a wide range of institutions, we recognize that differences in campus context may necessitate adaptations. For that reason, we begin with a short description of Carleton College, its students, and its economics curriculum.

Carleton is a highly selective liberal arts college located about one hour south of the metropolitan centers of Minneapolis and St. Paul. The institution is well-resourced with an average class size of 16 and a student-to-faculty ratio of 9:1. Of the roughly 2,000 students enrolled, approximately three-quarters graduated in the top decile of their high school class.<sup>2</sup> Another approximately 20 percent graduated in the top quintile. Students appear similarly strong when measured by board scores with about one-third earning SAT composite math + verbal scores of 1,500 or greater, while less than one-tenth earn scores below 1,200.<sup>3</sup> The student body is diverse along many dimensions. We draw essentially equal numbers of men and women and from almost every state in the union plus two dozen foreign countries. About one-quarter of students identify as students of color and one-eighth are among the first generation of their family to go to college. Just over half of Carleton students qualify for need-based Carleton grants.

The college offers no graduate degrees and steers clear of pre-professional programs. This institutional orientation is particularly relevant to the department of economics, because we offer no business degree. The economics major begins with the two-course sequence of Principles of Microeconomics and Principles of Macroeconomics. So, the two “Principles” courses play a foundational role in the department’s curriculum. Still, despite strong numbers of economics majors (just

---

<sup>2</sup> This statistic and those which follow are taken from the Class of 2020 Profile reported by Carleton’s admissions office: <https://apps.carleton.edu/admissions/apply/requirements/profile>.

<sup>3</sup> This statistic applies only to those students who chose to take the SAT.

under 10 percent of graduating seniors have completed a major in economics), the majority of students in any given section of Principles are destined to be non-majors and enroll in these courses to earn credit toward distribution requirements.

## Data

During the 2016-17 academic year, professors in all 16 sections of Principles of Microeconomics and Principles of Macroeconomics assigned students to complete the QLRA as an ungraded assessment. These courses drew a total of 374 student enrollments over three trimesters. Of these, 65 students enrolled in both a section of Microeconomic and Macroeconomic Principles and took the assessment twice. One re-enrolled in Principles of Microeconomics after a poor initial performance.

None of the teachers gave credit for either completion of or performance on the assessment. Students were told that their professors would review student scores in an effort to provide additional and better support to students whose quantitative reasoning background may be a bit weaker. Students were also given their own score and encouraged to use the information to inform their early use of academic support services. Because the assignment was not attached to any grading in the course, we were not surprised to see that some students chose not to complete the assessment. In all, of the 308 unique enrolled students, 219 (or 71 percent) completed the QLRA at least once.<sup>4</sup> Our unit of analysis is a student-course experience. Counting those who took both Principles courses and completed the QLRA, we have a total of 241 student observations in our sample.

Students completed the assessment online, outside of class time, and in an environment of their own choosing. We neither encouraged nor discouraged use of calculators. The instructions noted that Carleton students have completed the 20-item assessment “in about 30 minutes, though some take a bit more or less time.” No time limit was enforced. The mean completion time in our sample was 122 minutes with a median of 33 minutes. The large magnitude of the mean relative to the median reflects the behavior of a handful of students (18 in total) with durations longer than 3 hours. These students almost surely began the assessment and then left their browsers open to come back and complete the assignment at a later time. When these 18 students are excluded, the mean duration falls to 38 minutes, much closer to the median.

---

<sup>4</sup> We very intentionally decided not to assign explicit credit to completion of the assessment. Some of the professors involved were concerned that requiring a graded quantitative reasoning assessment early in the term might evoke math anxiety particularly among weaker students. These colleagues argued that our attempt to assess and support students could backfire by framing our courses in the context of mathematics performance anxiety. So, in an effort to “first do no harm,” as a group we agreed to a low-stakes administration.

While some students may have opted out by skipping the assignment altogether, we were also concerned that the low stakes might lead some to essentially opt out by clicking through the items without any real effort. The measure of duration suggests that this possibility was not a serious problem. Only six students spent less than 15 minutes completing the QLRA assessment. The mean score of these quick test-takers was 14.2 (out of a total of 20) as compared to a mean of 15.7 found among all others. This difference is not statistically significant ( $p = 0.30$ ) and so we assume that all students who completed the assessment did so with meaningful effort.

In light of the information above, we did not feel the need to discard any observations due to low effort. However, we were concerned about selection issues caused by the 29 percent of enrolled students who opted not to complete the assessment. To correct for selection, we stratified our sample by gender, class year, and grade earned in the course. (For the latter, we divided the data by full letter grades—those who earned A's, B's, C's, D's, and F's.) We then created probability weights equal to the inverse of the observed probability of response for each gender-class year-grade combination. For example, if only one-half of students of a particular type completed the QLRA, then each student who did complete the assessment is given a weight of 2 (i.e., the inverse of one-half) because each student observed by us represents two students in the population. Except where noted, all analyses below use these probability weights.

All of the other variables used in this study were drawn from institutional records. One complication arose in using ACT and SAT scores. While all students in our sample report either the ACT, the SAT, or both, we have neither ACT nor SAT scores for all students. For students without an SAT math score, we converted the ACT math score into an SAT math equivalent score utilizing reported percentile scores on both exams.<sup>5</sup> So, one can interpret the “SAT math” variable as a non-linear measure of the percentile score on available standardized tests.

Two final complications arose concerning our treatment of course grades. First, one student did not receive a grade by the time we collected our data. This student was removed from the analysis leaving us with 240 observations. Two other students dropped the course after the official drop-add deadline had passed. Such drops are published on the transcript and usually indicate that the student was headed toward a poor grade (or set of grades) and chose to drop one course to focus on others. Often students who late drop are headed toward grades in the dropped course in the D or F range. We did not want to eliminate these

---

<sup>5</sup> Percentile scores were retrieved from <https://www.act.org/content/dam/act/unsecured/documents/NormsChartMCandComposite-Web2015-16.pdf> and <https://collegereadiness.collegeboard.org/pdf/understanding-sat-scores-2016.pdf> on August 9, 2017. Exact crosswalk available from authors upon request.

observations, because they represent important forms of unsuccessful course completion. We coded these students' grades as 0.1 GPA points—something between an F+ (were there such a thing) and a D-. While our choice of this value will affect measures like ordinary least squares regression, our primary analyses are ordinal: so long as we are correct in coding these students' experiences as “unsuccessful,” the exact value we assign won't matter.

Table 1 provides descriptive statistics for the variables used in our analysis. Columns 1 and 2 note means and standard deviations using the probability weights while columns 3 and 4 report un-weighted counterparts. Note that the weighted and unweighted means are very similar. This observation suggests that students who opted out of the assessment did not have systematically high or low values for any of the variables. The weighted QLRA score might be slightly higher than the unweighted average, which suggests student groups with high QLRA scores are, on average, somewhat less likely to complete the assessment. Given that our purpose is to identify students who need extra support, this selection isn't terribly problematic for our purposes. Even this exception proves the rule: the weighted mean is only 0.17 higher, a magnitude that is just 0.05 standard deviations. Weighted and unweighted standard deviations are similarly alike. Columns 5 and 6 report minimum and maximum values for each variable.

**Table 1**  
**Descriptive Statistics of Independent and Dependent Variables**

Variable	Weighted		Unweighted		Minimum	Maximum
	Mean	St. Dev.	Mean	St. Dev.		
QLRA Score	15.82	3.44	15.67	3.51	2	20
SAT Math	721.22	61.85	719.92	60.84	510	800
Cumulative GPA*	3.44	0.39	3.44	0.37	1.39	4
Gender (1=male)	0.60	0.49	0.58	0.49	0	1
Class Year	1.40	0.81	1.41	0.81	1	4
Course Grade	3.16	0.77	3.13	0.76	0.1	4

Notes: Sample size = 240

\*Cumulative GPA was measured at the end of the 2016-17 academic year.

The sample is skewed toward men and first- and second-year students. This skew is expected in a gateway economics course. The means and standard deviations of the SAT math score and cumulative GPA are very representative of the Carleton population as a whole. Similarly, the course grade (averaging between a B and B+) is almost exactly average for our department in recent years.

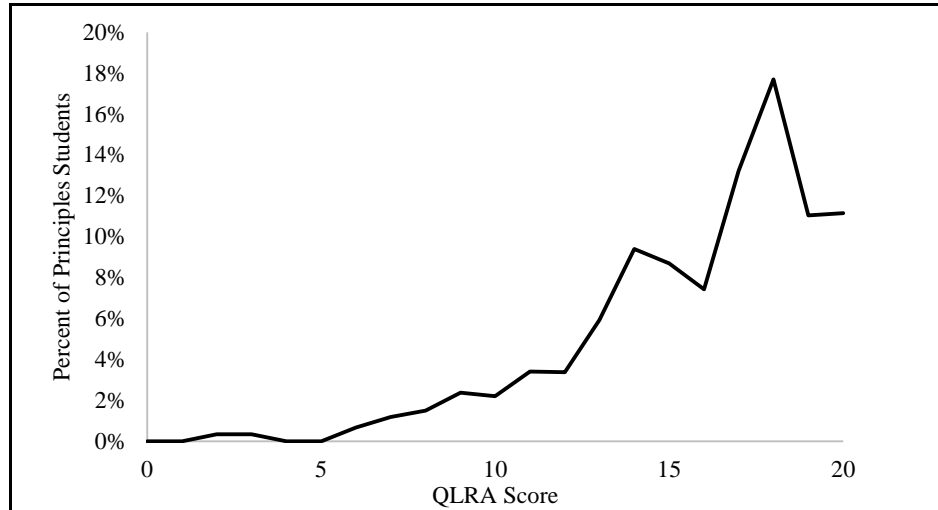
Figure 1 presents the distribution of QLRA scores with significant evidence of a ceiling effect: 11 percent of all students scored a perfect 20 out of 20 and another 11 percent correctly answered 19 questions. For some of the research questions explored below, this limitation biases the study against finding value in the QLRA. In particular, we want to know whether the QLRA provides any information not already embedded within ACT and SAT math board scores. To the extent that the signal of quantitative reasoning is muted at the top of the



QLRA distribution, we are less likely to find added value in the QLRA. However, our primary concern is with students at the bottom of the distribution where top-coding is not an issue.

## Analysis of Data

As we begin data analysis, we would again underscore the purpose of our study: to learn if and how the QLRA might be a useful formative assessment tool guiding student support in quantitative disciplines. Given that all of our data come from only one institution, and an outlier at that, we do not expect that the findings we present here will necessarily generalize to other contexts. Rather, we hope to share our process of thinking to show how others might similarly employ the QLRA for their own purposes in their own contexts.



**Figure 1. Distribution of QLRA scores.** Probability weights account for non-response.

The first question we explore is whether the QLRA adds anything in terms of our ability to predict course outcomes in Principles of Economics. After all, nearly all of our students have already taken the SAT or ACT. While it might be cumbersome to get that information out of institutional databases, if the QLRA added no new predictive power, then it may be preferable to spare our students the additional work of taking the QLRA. It may be particularly desirable to avoid additional work of taking the screening test with students who may, as a result of the assessment, associate economics with the negative feelings of math anxiety.

To assess the QLRA's "value added," we first ran an ordinary least squares regression predicting course grade measured on a four-point scale with half-grades (e.g., A- versus A) worth 0.33 grade points. In addition to scores from the QLRA and the SAT math exam, we included variables a professor might have

readily on hand: gender, class year, and overall academic performance in college as measured by the cumulative GPA.<sup>6</sup> The first column of Table 2 presents results. Not surprisingly, scores on the QLRA and SAT math along with general

**Table 2**  
**Estimating Course Performance Using Full Set of Independent Variables**

<b>Outcome Analysis</b>	<b>Course Grade OLS (1)</b>	<b>Course Grade Tobit (2)</b>	<b>B or Better Logit (3)</b>	<b>C or Better Logit (4)</b>
<b>QLRA Score</b>	0.065 (0.016)	0.081 (0.019)	0.014 (0.007)	0.015 (0.005)
<b>SAT Math</b>	0.003 (0.001)	0.003 (0.001)	0.001 (0.004)	0.000 (0.000)
<b>Cumulative GPA</b>	0.774 (0.089)	0.973 (0.122)	0.311 (0.0592)	0.067 (0.033)
<b>Gender (1=male)</b>	0.063 (0.078)	0.071 (0.101)	0.103 (0.046)	0.008 (0.031)
<b>Class Year</b>	0.015 (0.038)	0.034 (0.054)	0.017 (0.028)	-0.004 (0.012)
<b>Constant</b>	-2.610 (0.485)	-3.670 (0.629)		
<b>Observations</b>	240	240	240	240

*Notes: Standard errors are reported in parentheses. In models 1 and 2, the dependent variable is the grade earned in the course. The reported coefficients can be interpreted as the average change in grade predicted due to a one-unit change in the independent variable (other variables held constant). In models 3 and 4, the dependent variable takes on a value of 1 or 0 depending on whether the student achieves or fails to achieve the relevant level of success. Reported coefficients reflect the estimated change in probability of success due to a one-unit change in the independent variable (other variables held constant) estimated at mean values.*

academic performance all positively predict the ultimate course grade. Important for our purposes, the QLRA has predictive power independent of the SAT and other course grades. Controlling for all other independent variables, a one-standard deviation increase in QLRA (about 3.44 points on the 20-point instrument) predicts a course grade that is higher by 0.21 grade points—not quite one-half letter grade. This is a slightly larger impact than that of a one-standard deviation increase in SAT Math score. After controlling for these performance measures, demographic markers of gender and class year show no predictive power.<sup>7</sup>

<sup>6</sup> Technically, the professor would not have our measure of cumulative GPA because it was captured at the end of the 2016-17 academic year. Indeed, many of our fall term students were in their first term and so a professor could have no prior information on academic performance.

<sup>7</sup> We repeated the analysis here including an indicator variable noting whether the student was from a low-income family and/or was among the first generation of her family to attend a 4-year college (LIFG). In no model was this variable statistically significant—either in its level or when interacted with the QLRA score. Having controlled for performance in the form of SAT math,

While ordinary least squares is a common starting point, it is inappropriate for our purposes due to censoring at the top of the grade distribution. In our sample, 46 of 240 grades recorded were the maximum grade of A (a numerical value of 4.0). Though we might expect that some of these students might be stronger than others, variation in performance cannot be reflected in variation in course grade once the maximum is reached. In statistical terms, we have censored data with a reported grade of A meaning that a student performed at a level that is “at least” a 4.0. It is easy to visualize the consequences of this censoring in the data. We begin with a cloud of data with an upward slope. The censoring then pushes the top-most observations down to some pre-determined maximum score. Because there is a positive correlation between our two imagined variables, the top-most data are also the right-most data. Thus, the censoring affects the right side of the scatter plot more than the left and the estimated slope is biased toward zero.

Column 2 of Table 2 reports the results of a Tobit regression that accounts for the censoring of grades at the upper level of 4.0. Correcting for censoring does not change any of the qualitative results. However, the magnitude of the QLRA coefficient is about one-third larger. Following the Tobit correction, a one-standard deviation change in QLRA performance is associated with a grade improvement of 0.28 grade points, notably larger than the effect of a similar one-standard deviation change in SAT Math score.

While these results are encouraging, they are not particularly focused on our purpose. Our primary goal is early identification of students who are headed toward an unsuccessful experience so that we can provide targeted intervention. This question is more concerned with the bottom of the grade distribution than the top. Columns 3 and 4 of Table 2 analyze dichotomous metrics of success defined by earning a B or better (column 3) or a C or better (column 4). Using logistic regression (or “logit”) analysis we can estimate the change in probability of clearing a given threshold conditional on other independent variables.

Results in the two logit models continue to point to important, independent predictive power of the QLRA. Controlling for the other factors, at mean levels of variables, a one-standard deviation increase in the QLRA is associated with a 5-percentage point increase in the probability of success at either the B or C level. Given that over 70 percent of students currently earn a B or better and more than

---

QLRA, and cumulative GPA, LIFG students showed no difference in performance. Of course, this does not mean that LIFG students don't need disproportionate support. As discussed above, in our sample, LIFG students had lower SAT math scores, QLRA scores, and cumulative GPAs. As representation of LIFG students increases as campuses diversify, it becomes more important that our teaching practices and pathways to majors are inclusive of students that have had less rigorous coursework prior to college. For, as Engstrom and Tinto put it so aptly, “access without support is not opportunity” (2008).

90 percent achieve at least a C, a 5-point increase in the probability would represent a substantial portion of students who fall short of these targets. Moreover, it is notable that a one-unit increase in QLRA score predicts the same increase in probability of success at the C level as at the B level. This result is in contrast with SAT Math and cumulative GPA which are less associated with success at lower levels of performance. This finding is important because we are particularly interested in targeting those who are headed toward low grades.

Because the logit model is non-linear, the effect of a one-unit change in an independent variable does not produce the same effect size at all values of the variable. If the QLRA score were to increase from 5 to 6 correct answers, the resulting change in probability of earning at least a B increases by 2 percentage points.<sup>8</sup> This is twice the change in probability associated with an increase in QLRA score from 19 to 20. When success is measured at the C level, the impact of a QLRA increase from 5 to 6 is four times as large as that resulting from a QLRA increase from 19 to 20. In other words, the QLRA is particularly predictive at low levels. This bodes well for using it as a mechanism for targeting academic support.

The results in Table 2 show that the QLRA provides information above and beyond that already known to the institution. Of course, if a professor or department were interested in using the QLRA to target support, it might be more realistic to expect limited access to institutional data like SAT scores or cumulative GPAs. So, we next consider using the QLRA alone to screen students for additional academic support. In such diagnostic work, we worry about two types of errors. A “false positive” occurs when the QLRA singles out a student for additional support even though that student would succeed even in the absence of such assistance. In this case, intervention ties up resources that might have been devoted to other students with greater need and may waste the student’s time. A “false negative” occurs when a student scores well on the QLRA and yet is headed toward an unsuccessful outcome. In such a case, even though the student has sufficient quantitative reasoning skill, other skill gaps may inhibit success. An ideal metric would direct support to all students headed toward an academic road bump and only to such students.

Of course, the rates of false positive depend on both our definition of course success and the cutoff in QLRA scores used to identify students in need of additional support. Raising the grade threshold for success results in more and more true positives because more and more students would fall below the bar. On the other hand, if the QLRA cutoff is increased we would miss fewer failing

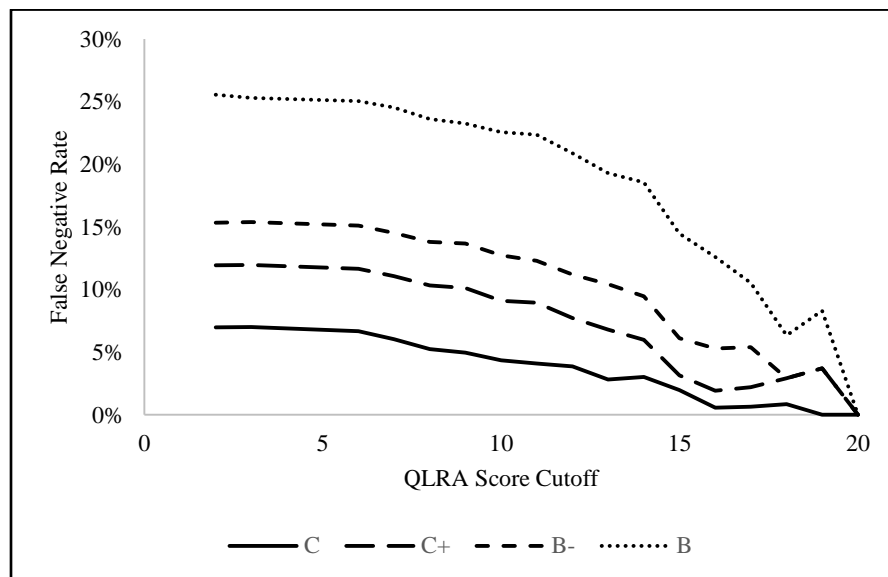
---

<sup>8</sup> This and the other results reported in this paragraph must be computed using the non-linear probit model and are not directly computable from the results presented here. They were reported by the statistical package (Stata 11.0) used to estimate the model.

students but end up intervening with more and more students who would have been fine without our assistance.

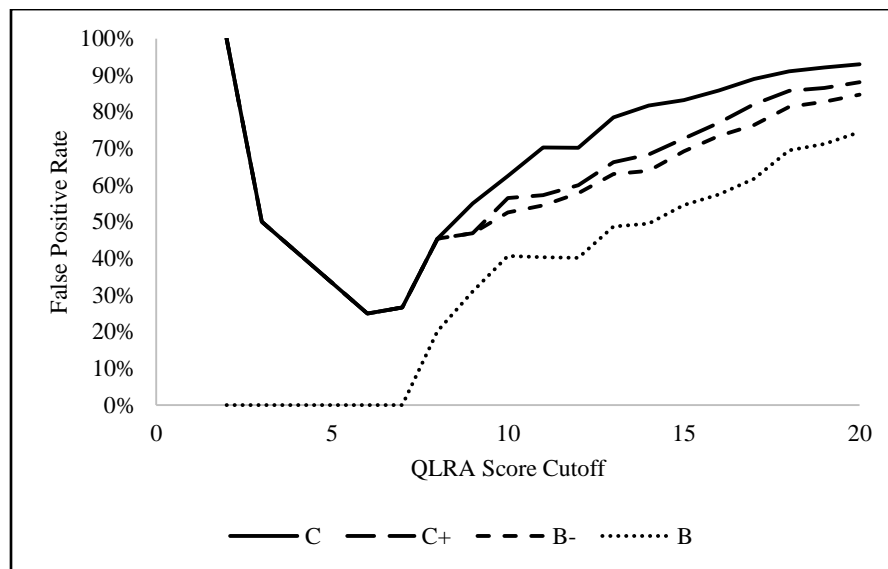
Figures 2 and 3 plot the rates of false negatives and false positives, respectively, that would have resulted from every possible definition of success from C- to A and every cutoff for academic support from a QLRA score of 2 to 20. By construction, we face a tradeoff between the two error types. In addition to this balance between statistical errors we are cognizant of resource constraints that limit our ability to intervene meaningfully with large numbers of students.

In Figure 2, two patterns speak positively to the value of the QLRA as a screening device for our courses. First, the false negative rate is relatively low: in our context, a student who is headed toward a weak grade in Principles of Economics nearly always posts a low QLRA score. For example, if we define B- or better as a successful experience in the gateway course, even a very low cutoff in the QLRA would miss only one-in-seven students on a path toward a sub-par grade. Second, the rate of false negatives doesn't change much across reasonable cutoffs in the QLRA. This characteristic allows us to pay more attention to the false positive rate when setting the cutoff.



**Figure 2. False negative rate as a function of QLRA cutoff for definitions of success ranging from a course grade of C to B.** A false negative occurs when the QLRA score indicates that a student will succeed without assistance even though the student will in fact fail to reach the desired level of success. For example, if we provide assistance only to students scoring 10 or lower on the QLRA and “success” is defined as a B- or better, then approximately 15 percent of students heading toward “low” grade will not be offered additional help.

The false positives reported in Figure 3 are notably higher. There are two reasons why it is not clear that these higher rates of false positives are necessarily a significant problem. First, it may be an artifact of the low-stakes environment of our QLRA test. If students knew that we would expect low QLRA scorers to attend some academic support events, then they might put more effort into the QLRA and that would result in fewer false positives. Second, assuming the score resulted from reasonable effort, we would not regret providing services in support of QR to students who need them even if they would have earned a reasonable grade in Principles of Economics without the support. Still, false positive rates at or above 50 percent mean resources diverted away from the subgroup of most immediate concern.



**Figure 3. False positive rate as a function of QLRA cutoff for definitions of success ranging from a course grade of C to B.** A false positive occurs when the QLRA score indicates that a student is at risk for failure and needs assistance even though the student will succeed without help. For example, if we provide assistance to all students scoring 10 or lower on the QLRA and “success” is defined as a B or better, then approximately 40 percent of students receiving assistance would have succeeded on their own without additional help.

## Concluding Discussion

The results presented in the previous section suggest that the QLRA may be a low-cost screening tool for an economics gateway course, identifying students whose quantitative background may portend poor course performance. While our data speak directly only to the context of economics at our own institution, we believe other quantitative disciplines and other institutions might find the QLRA useful in similar ways.

Of course, the next step is more difficult. Having found a tool for quickly identifying students in need of additional support, next we need to develop interventions that help these students achieve success. Recognizing that the math anxieties and skills gaps that underlie the QLRA results have developed over more than a decade of full-time education, we must be humble in our approach. Realistically, we have about three weeks between identifying a struggling student and the first midterm. So, we do not expect to raise students from the bottom of the quantitative reasoning skill distribution to the top. Still, the findings in the previous section suggest that if we can move the needle even a little, we can noticeably increase the probability of success.

Informal discussions have identified several possible paths forward. First, Carleton colleagues in other departments offer sections of gateway courses “with problem solving.” These courses augment typical class meetings three days per week with one or two more meetings in which students work in teams to solve problem sets. These sessions might be thought of as “group office hours,” with the entire class engaged in active learning with the professor’s intermittent intervention. The additional contact time also allows for some just-in-time instruction to refresh relevant high school math concepts. By forcing group work and active-learning strategies, our colleagues report noticeable improvements in the performance of students who come into the course with skills gaps. These colleagues underscore that the “with problem solving” sections service the same middle-level courses in their majors, and so these sections must end up covering the very same content as other versions of the gateway course.

We have also been in conversation with student support services about possible workshops designed to refresh relevant math skills and reduce students’ math anxieties. Given the heavy workload that students already carry, we recognize that any add-on of this sort must be modest in size to avoid placing counterproductive burdens on student time.

Finally, in collaboration with the Liberal Arts Consortium for online Learning and the Associated Colleges of the Midwest we have been developing a suite of online tools (Qbits) designed to provide targeted, just-in-time instruction on basic math concepts. For example, we are building modules around linear equations, log functions, and solving linear systems of equations. Each Qbit will contain pre-testing, a brief instructional video, applications drawn from a range of disciplines, and post-testing. Operated through our course management system, Qbits can be assigned by professors to weaker students precisely at the time when the course will use the mathematical concepts.

We are sure many more ideas will arise through discussion (and from articles published in this journal).

## Acknowledgments

We are thankful to Shane Auerbach, Theresa Bauer, Faress Bhuiyan, Bruce Dalgaard, Stephie Fried, Ben Keefer, Jonathan Lafky, and Aaron Swoboda for assigning the QLRA to their students.

## References

- Asai, David J., and Cynthia Bauerle. 2016. "From HHMI: Doubling Down on Diversity." *CBE-Life Sciences Education* 15 (3 fe6): 1-3.
- Association of American Colleges and Universities. 2007. *College Learning for the New Global Century*. Washington, DC: Association of American Colleges and Universities.
- Bailey, Thomas, Dong Wook Jeong, and Sung-Woo Cho. 2010. "Referral, Enrollment, and Completion in Developmental Education Sequences in Community Colleges." *Economics of Education Review* 29 (2): 255–270.
- Engstrom, Cathy, and Vincent Tinto. 2008. "Access Without Support is Not Opportunity." *Change: The Magazine of Higher Learning* 40 (1): 46–50.
- Gaze, Eric C., Aaron Montgomery, Semra Kilic-Bahi, Deann Leoni, Linda Misener, and Corrine Taylor. 2014. "Towards Developing a Quantitative Literacy/Reasoning Assessment Instrument." *Numeracy* 7 (2): Article 4. <https://doi.org/10.5038/1936-4660.7.2.4>
- Hacker, Andrew. 2012. "Is Algebra Necessary?" *New York Times*, July 28.
- Kuh, Georges D., Brian K. Bridges, and John C. Hayek. 2006. *What Matters to Student Success: A Review of the Literature* (Vol. 8). Washington, DC: National Postsecondary Education Cooperative.
- National Center for Education Statistics. 2016. *Digest of Education Statistics*. Washington, DC: U.S. Department of Education.
- Vacher, H. L. 2011. "A LEAP Forward for Quantitative Literacy." *Numeracy* 4 (2): Article 1. <https://doi.org/10.5038/1936-4660.4.2.1>
- Wang, Ming-Te, Jessica Degol, and Feifei Ye. 2015. "Math Achievement is Important, but Task Values are Critical, Too: Examining the Intellectual and Motivational Factors Leading to Gender Disparities in STEM Careers." *Frontiers in Psychology* 6 (36). <https://doi.org/10.3389/fpsyg.2015.00036>
- Warburton, Edward C., Rosio Bugarin, and Anne-Marie Nunez. 2001. *Bridging the Gap: Academic Preparation and Postsecondary Success of First-Generation Students. Statistical Analysis Report*. Washington, DC: National Center for Education Statistics.