# An Assessment Study of the Effectiveness of the General Education Quantitative Reasoning "B" Requirement at the University of Wisconsin-Madison

*A Report Submitted to the University General Education Committee*


Charles N. Halaby
Research Director
Associate Dean, College of Letters and Science

General Education Research Group

Nancy Westphal-Johnson
Chair, University General Education Committee
Associate Dean, College of Letters and Science

Char Tortorice
Director, Testing and Evaluation
Wisconsin Educational Research Center

Elaine Klein
Assistant Dean, College of Letters and Science


November, 2006

# An Assessment Study of the General Education Quantitative Reasoning "B" Requirement at the University of Wisconsin-Madison

## 1.0 Introduction

*Quantitative Reasoning A and B* are two key provisions of the UW-Madison General Education Requirements that were adopted by the Faculty Senate in 1994 and implemented in 1996. Over the past decade, the implementation of these requirements has been accompanied by both ongoing and periodic special-purpose studies aimed at the systematic assessment of course, curricular, and student outcomes.[1] Last year we reported the results of three related assessment studies of student outcomes achieved during 2004-05 under part "A" of the Quantitative Reasoning (hereafter, QR–A) requirement.[2] That research was the first to use measures other than course grades to assess directly the extent to which satisfying a quantitative reasoning requirement yields student outcomes that are consistent with general education objectives. The current report complements the earlier one insofar as it describes the results of an assessment of student outcomes achieved during 2005-06 under the Quantitative Reasoning "B" (hereafter, QR–B) requirement. The primary goal is to assess the extent to which courses that are certified as satisfying the QR–B requirement yield student outcomes that fulfill fundamental general education objectives with respect to quantitative reasoning proficiency.

The aim of the QR–B requirement is to enhance quantitative proficiency by exposing students to courses in which, in the context of other substantive material, the application of quantitative methods

[1] The first assessment undertaken specifically in connection with the QR requirements was carried out on an ongoing basis from 1994-2004 by the Quantitative Assessment Project, a program of course-specific evaluations that was started in 1990 and later was adapted to the QR objectives. Quantitative Assessment Project (QAP) has helped faculty members by creating examinations to measure students' mathematical skills in areas most germane to topics covered in specific classes. The project is described in a 2003 report by Alvarez-Adem and Robbin that may be found at http://www.ls.wisc.edu/Gened/FacStaff/QRAssessReport20022003.pdf. In 2001 a subcommittee of the L&S General Education Committee carried out a more broad ranging curricular-level assessment aimed at identifying the course-taking and performance characteristics of students as they pursued fulfillment of the QR requirements. A description of this assessment effort, together with other General Education assessment initiatives, may be found at http://www.ls.wisc.edu/Gened/FacStaff/assessment.html.

[2] See *Two Assessment Studies of the General Education Quantitative Reasoning "A" Requirement at the University of Wisconsin-Madison* at http://www.ls.wisc.edu/Gened/Assessment/default.htm

and concepts is a central feature and theme. The courses that fulfill the QR–B requirement span a vast variety of content spread across a wide range of disciplines, literally from A (Astronomy) to Z (Zoology), but all are expected to satisfy a common set of criteria. As described by Professor Richard Brualdi in a 1994 founding document, a QR–B course must significantly involve:

- the recognition and construction of mathematical models and/or hypotheses that represent quantitative information
- the evaluation of these models and hypotheses
- the analysis and manipulation of mathematical models
- the drawing of logical conclusions, predictions or inferences
- the assessment of the reasonableness of conclusions

These criteria constitute what may be thought of as a core set of dimensions of quantitative reasoning along which both courses and students may be assessed. The fundamental objective of the research reported here was to determine the extent to which courses that satisfy the QR–B requirement, and hence putatively fit these criteria, provide students with the experience and opportunity needed to enhance their quantitative reasoning skills. In other words, the goal was to fashion one study that could serve as an assessment of both QR–B courses and student outcomes.

To anticipate, our research method amounted to asking a sample of students *how much* the courses and educational experiences they had had during the semester just completed had taught them the skills that the QR–B requirement views as critical to achieving quantitative proficiency. The heart of the analysis we present involves rigorously comparing students who had and had not taken a QR–B course in terms of their self-reported gains in quantitative reasoning proficiency along various skill dimensions. The results of such a comparison may be viewed from two vantage points. First, such comparisons reflect directly on the differences in the content, as experienced by students, of courses that do and do not satisfy the QR–B requirement. The general education objectives anticipate that students who satisfy the QR–B requirement by taking courses that have been certified as delivering the appropriate content would have a richer quantitative reasoning experience than they might otherwise, and richer than the experience of students who do not take a QR–B course. This study will assess the extent to which that belief is empirically valid.

Second, to the extent student self-reports of gains in quantitative reasoning proficiency may be viewed as valid, if indirect, indicators of actual student learning outcomes, the comparisons we construct can yield estimates of the effect of the QR–B requirement on quantitative reasoning skills. The 2004-05 study of the QR–A requirement provided strong evidence of a close connection between estimates of gains in quantitative reasoning proficiency yielded by student self-reports and estimates yielded by laboratory tests of actual student learning. An comparison of students who did and did not satisfy the QR–A requirement in terms of their self-assessed gains in quantitative reasoning yielded results that were closely corroborated by rigourously estimated differences in quantitative reasoning ability as measured by laboratory tests. These earlier results buttress the argument for the validity of self assessment as a tool for the comparative analysis of learning outcomes, at least in the context of quantitative reasoning skills. Hence, we believe there is ample ground for interpreting the analysis reported below both in terms of differences in the content of QR–B as compared to other courses, as well as in terms of differences in student learning outcomes.

The remainder of this report is organized as follows. The next section describes the sampling procedure, sources of sample selection bias, and the measurement of improvement in quantitative reasoning proficiency. Sections 3 and 4 present the analysis and conclusions, respectively.

## 2. Data and Measurement

### 2.1 *Target Population and Sampling*

The assessment issues driving this research were examined with a sample of students who were enrolled fall semester 2005-06. Since the main goal was to compare the learning experiences of students who did and did not take a QR–B course in a given semester, a stratified sampling design was appropriate and efficient.[3] To that end, we first identified the population of QR–B certified courses that would be the basis for establishing a sampling frame for selecting students. There are some 75 courses at UW-Madison that are listed as satisfying the QR–B requirement, but only 39 of these were offered in fall 2005. Another 14 courses that were explicitly mathematical or statistical in nature, or explicitly required calculus as a prerequisite, were also eliminated because their quantitative reasoning content appeared on its face to surpass general education objectives.[4] That left 25 courses whose fall rosters were the starting point for defining our population of "QR–B" students.

I say starting point because we hoped to compare students who were taking their first QR–B course during fall 2005 to students who had not yet fulfilled the QR–B requirement. To that end, we chose to focus on students who were either freshmen or sophomores, and thus still early in their UW-Madison careers; this is also the time when the bulk of students fulfill the QR–B requirement. These additional restrictions reduced the QR–B sampling frame to 1486 students who were enrolled in their first QR–B course during fall 2005-06. Applying similar restrictions to the population of courses that did not satisfy the QR–B requirement yielded a sample of 1,398 freshmen and sophomores who, at the completion of fall semester 2005, had never taken a QR–B course at UW-Madison. Samples of 160 students were randomly selected from each of these two population strata.

Judging from table 2.1, there appear to have been minor inaccuracies in either the population

[3] By "efficient" I mean this design would be expected to yield smaller sampling error in the estimates of the effect of QR–B than would a simple random sample of the same size.

[4] Deleting these courses was also justified by the desire to concentrate the sample on students who were believed to be a primary "target" of the QR–B requirement, namely, those who might not otherwise develop the desired level of quantitative proficiency. By the same token, introductory statistics courses within substantive disciplines like sociology and psychology were not eliminated from the target stratum of QR–B courses.

database or the construction of the sampling frame with respect to students' QR–B status prior to fall 2005. The rows of this table indicate whether a student had taken a QR–B course prior to fall 2005; the marginal distribution indicates that 19 of the 320 students who were selected for the sample had previous exposure to QR–B. Because they represented errors in the sampling frame, these nineteen students were eliminated from the sample, thereby leaving 301 selected cases.

The other sampling restriction was year of matriculation. Among the 301 students who had not previously taken a QR–B course, 59% were sophomores and 36% were freshman, with the remaining 5% having matriculated in 2003-04 or earlier. Like the 19 students who had previously taken a QR–B course, students who had matriculated before 2004-05 represent errors in the sampling frame; eliminating them yields a final sample size of 285. Of these 285 students, 161 (56%) had enrolled and 124 (46%) had not enrolled in a QR–B certified course during fall 2005-06.

### 2.2 *Sample Selection Bias*

The figures just reviewed pertain to students who, having been selected for the sample, were sent a mail survey to fill out and return.[5] Of the 285 students who were members of the valid sampling frame and invited to participate in the study, 214 (75%) actually returned a usable questionnaire and constitute the basis for the results reported here.[6]

An important consideration is whether the eventual respondent sample is relatively representative of the target population or subject to systematic bias. This question can be examined by considering the association between student characteristics and the propensity to participate in the survey. One key issue is whether response rates depended on QR–B status? Were students who were enrolled in a QR–B course more eager to fill out and return a questionnaire devoted to quantitative reasoning

[5]The survey was conducted by the University of Wisconsin Survey Center under the direction of Bob Craddock. The survey went in the field early spring semester 2005-06.

[6]If we had used the invalid sampling frame and included in the sample respondents who failed to meet our sample criteria the response would be about the same, 73%.

skills? Table 2.2 gives the target sample sizes, actual sample sizes, and response rates for the two population strata. The last column of this table shows that students enrolled in QR–B were slightly more likely to respond to the survey, although the three percentage point difference is far from being statistically significant ($\chi^2$ = 0.34; p-value= 0.56). If there were a selection bias associated with enrollment in QR–B courses, one would expect that students who took more QR–B courses would be even more likely to participate in the survey. Table 2.3 speaks to this hypothesis by giving the rates of survey response by the number of QR–B courses taken in fall 2005-06. As this table shows, rate of response does not increase systematically with number of QR–B courses, for students with one or more QR–B courses are about equally likely to have responded to the survey; again, there is no statistically significant difference in response rates ($\chi^2$ = 1.07; p-value=.78). Overall, then, the propensity of students to participate in this study, given their selection into the sample, does not depend on their QR–B status.[7]

The most serious potential source of sample selection bias would occur if the propensity to respond was a function of student characteristics that might be related the dependent variable of this study, quantitative reasoning proficiency. Two such sources that suggest themselves and are of some interest interest in their own right are gender and academic performance. As Table 2.4 shows, females were slightly more likely to return a survey than their male counterparts, although the 4% point difference between the female rate (76.7%) and the male rate (72.48%) is far from statistically significant ($\chi^2$ = 0.96; p-value=.42). With regard to academic performance, stronger students tended to respond to the survey at a higher rate than their academically weaker counterparts. Hence students whose fall grade point average exceeded the median of 3.1 responded at a rate of 81%, while the response rate was only 69% among students whose overall first semester GPA fell below the median. This twelve percentage point difference, which represents the strongest association we have found with respect to

[7] A similar result was found in the 2004-05 study of the QR–A requirement: students who did and did not take a course to satisfy the requirement responded to the survey at the same rate.

survey participation, is statistically significant (t=2.38; p-value=.018).[8]

Interestingly enough, although academically stronger students were more likely to respond to the survey, performance in QR–B courses was not itself directly associated with participation. In last year's QR–A study, QR–A course performance was positively associated with survey response, which meant that the sample of QR–A students represented not the average QR–A student, but rather those who had received higher grades in their QR–A course. This kind of relationship did not extend to performance in QR–B course: among students who satisfied the requirement in fall 2005, those who did well in their QR–B course were no more likely to respond to our survey than students who did less well. Aggregating across the grades of students who took 1-3 QR–B courses, the association between grade and response rate is extremely weak statistically($\chi^2 = 5.02$; p-value=.414). Hence, what we learn below about QR–B courses, their content and their effect on reasoning skills, does not reflect the biased representation of students who did well in their QR–B courses, since such performance is not associated with selection into the sample.

In summary, the only characteristic that appears to be associated with survey response and therefore needs to be accounted for by the analysis is fall semester academic performance.

### 2.3 *Measurement of Quantitative Reasoning Content and Skills*

The outcome variable for this assessment was measured by a scale constructed from a set of Likert-type items that tapped students' self-reported gains in quantitative reasoning skills achieved during fall semester 2005. Most of the items are similar to those used in last year's QR–A survey.[9] In order to encourage all students to reflect broadly upon their learning experiences during fall semester 2005,

[8] There was no association at all between scores students achieved on the mathematical placement test and their propensity to participate in the survey.

[9] A preliminary set of items intended to capture the various dimensions of quantitative reasoning was developed by visiting web sites that addressed the content of the concept as defined at colleges and universities with general education requirements. This set of items was then revised in consultation with Professor Richard Brualdi of the Department of Mathematics, who was instrumental in the development of the QR requirements and is in charge of approving new QR course proposals in his capacity as the Quantitative Reasoning liaison for the campus.

the preamble to the main question read as follows:

> *Please think about all the courses that you took here at UW during the fall semester. This includes the lectures, readings, and assignments, as well as studying with other students and any other experiences you had as part of these courses.*

This preamble was followed by the question:

> *How much did these courses and educational experiences teach you to do each of the following*:

where the items used to measure the quantitative reasoning content of courses and the skills students putatively gained from them during fall semester were listed as follows:

- Recognize logically sound arguments
- Express ideas using quantitative information
- Understand the difference between correlation and causation
- Use statistics to evaluate factual claims
- Use mathematical models to express ideas
- Recognize when arguments use evidence well
- Use quantitative information to solve problems
- Know when it is valid to infer that one thing causes another
- Use quantitative information to evaluate an argument
- Manipulate mathematical models to draw conclusions
- Understand models and hypotheses that represent quantitative information

with response categories ranging from "not at all', scored 1, to "a great deal", scored 5. Averaging over these eleven items for each respondent yielded a scale (hereafter, the "general" QR scale) with Cronbach's $\alpha$–reliability equal to .87, which is excellent. The initial analysis reported below assumes that this scale is unidimensional in the sense that the constituent items represent a single underlying concept. Later the scale will be disaggregated into subscales to gauge the extent to which satisfying the QR–B requirement may have differential implications for constituent dimensions of quantitative reasoning.

## 3. Analysis and Findings

### 3.1 *How the QR–B Requirement Affects Self-Reported Gains in Quantitative Reasoning Proficiency*

This section compares students who did and did not take a QR–B course with respect to the general quantitative reasoning content of their educational experiences during fall 2005-06. Table 3.1 gives an initial set of results generated by regressing the general QR scale on indicators of QR–B course experience and salient control variables. The first column gives the results for a model that simply contrasts the mean scale scores of students who did and did not take QR–B certified courses. The result indicates that the mean scale score for QR–B students (3.47) exceeded the mean scale score for other students (3.29) by .179, which is statistically significant at the .08 level.[10]

This result, which suggests that fulfilling the QR–B requirement improves quantitative reasoning skills, is not as statistically weak as it might appear at first blush. Remember that, first, students are not being asked about the QR content of QR–B courses relative to all other courses they have taken; and second, for most students a QR–B course constitutes a small fraction of their total fall semester educational experience. If this observation has merit, we should find that as QR–B courses become a larger part of a students experience, differences in quantitative reasoning gains picked up by our scale should increase. Table 3.2 and Figure 1 show that this is exactly what happens: As the number of QR–B courses increases, the higher the average score on the general QR scale. In addition, because the pattern of increase in Figure 1 follows a straight line, a linear model suffices to capture the effect of QR–B courses on quantitative reasoning. This model is given in column 2 of table 3.1, according

[10]For models 1 through 3, the row labelled "constant" gives the mean QR scale score for students who did not satisfy the QR–B requirement.

to which each additional QR–B course yields a .147 (t-ratio=3.00; p-value=.01) mean increase in the general QR scale.

The validity of our quantitative reasoning scale is buttressed by its association with the number of QR–B courses a students takes, since we would expect more courses to yield greater self-reported gains in quantitative reasoning skills. Extending this same logic suggests that students who perform better in their QR–B courses should also report higher average scores on the QR scale if the latter is indirectly tapping improvement in student skills and not mere exposure to quantitative reasoning course content. In order to model this claim, the effect of QR–B on student quantitative reasoning may be conceptualized as having two components, one that is common to all students who enroll in QR–B, and a second component that reflects each student's QR–B performance. A model for this idea can be written as:

$$y_i = \alpha + \beta Q_i + \lambda P_i + \epsilon_i \tag{1}$$

where $Q_i$ is a variable that indicates the number of QR–B courses a students took in the fall, and $P_i$ is academic performance as measured by QR–B grade point, with students who did not take a QR–B course scored 0 on both variables. Given this coding, the QR scale mean for students who did not take QR–B is given by the intercept, $\alpha$. The expected value of the QR scale score for the typical QR–B student may be calculated as:

$$(\alpha + \beta\mu_1 + \lambda\mu_2) \tag{2}$$

where $\mu_1$ is the average number of QR–B courses taken by QR–B students, and $\mu_2$ is the average grade in QR–B courses.[11] The average effect of QR–B certified courses on quantitative reasoning skills is then the difference between these two parameters, namely, $(\beta\mu_1 + \lambda\mu_2)$.

[11] The means $\mu_1$ and $\mu_2$ were estimated by their sample counterparts. For fall semester 2005-06, the mean number of QR–B courses for students who took QR–B at all was 1.43, and the average grade across all QR–B courses was 2.78. In the fall, there was no gender difference in mean QR–B grade: males averaged 2.27 and females averaged 2.23.

Estimates of the parameters of this model appear in column (3) of Table 3.1. These estimates may be used to calculate an average effect of QR–B certified courses on quantitative reasoning skill across all QR–B students; or they can be used to calculate an average effect for students who took just one QR–B course. In the first case, we estimate that QR–B certified courses increase the QR scale mean by .32 units, which is highly statistically significant (t-ratio=3.47; p-value=.001). Among students who take only one QR–B course, the estimated QR gain is slightly less, .28, but this too is statistically significant (t-ratio=3.28; p-value=.01).

The remaining model of Table 3.1 controls for overall fall semester grade point in order to assess the robustness of the results to this point. The findings in column 4 indicate that QR–B courses continue to yield significant increases in self-reported QR gains after taking account of overall academic performance. In this model, the contribution of QR–B to quantitative reasoning is appropriately evaluated by testing the hypothesis that both coefficients of courses and QR–B grade are simultaneously zero. This hypothesis is rejected: Together, the coefficient of QR–B courses and QR–B grade are jointly statistically significant (F=4.63 p=.01). Computations analogous to those reported above yields an effect of QR–B certified courses across all students of .28 (t-ratio=2.89; p-value=.004) and across students with only one QR–B course of .23 (t-ratio=2.58;p-value=.01), both of which are statistically significant. Note that model 4 also shows that students of higher academic achievement as indicated by the overall grade point average tended to report greater gains in quantitative reasoning skills, a finding that is significant at the .08 level. This last result is interesting in its own right because it appears even with the number of and performance in QR–B certified courses held constant.[12]

### 3.2 *How the QR–B Requirement Affects Self-Reported Gains on Specific Dimensions of Quantitative Reasoning*

[12] We also fit a model that controlled for whether a student was QR–A exempt and controlled for gender. Neither QR–A exempt status (t=0.52) nor gender (t=0.73) are related to the general QR scale, and hence none of the conclusions based on models 3 and 4 are affected.

To this point the analysis has been carried out exclusively in terms of a single summative scale consisting of eleven items that target self-reported improvements in quantitative reasoning skills during first semester 2005-06. We have assumed implicitly that "quantitative reasoning" is a unidimensional construct: factors that cause one aspect of the underlying phenomenon tapped by the construct not only have the same direction of effect, but the same magnitude of effect, on *all* aspects of the underlying phenomenon. Yet this assumption clashes on its face with what is known about the quantitative reasoning phenomenon, and what is known about the content of quantitative reasoning courses. With regard to the phenomenon, the scale items themselves are quite heterogeneous, with some referring to more purely mathematical concepts and operations, and others referring to aspects of reasoning that pertain to logic, argument, and causation. Similarly, courses vary with respect to the quality and quantity of their quantitative reasoning content. Finally, last year's study of QR–A showed, using a similar scale, that satisfying the QR–A requirement had radically disparate effects on the different aspects of quantitative reasoning tapped by the overall scale.

The analysis that follows looks more closely at the effects of fulfilling the QR–B requirement on ostensibly different dimensions of quantitative reasoning skill. To this end, we disaggregated the QR–scale into four subscales that an exploratory common factor analysis revealed to be a reasonably good representation of the constructs underlying the 11 items; 1 item stood alone as its own scale. The scale labels, their $\alpha$–reliability, and their constituent items are as follows.

- **Arguments**(.70)
  - Recognize logically sound arguments
  - Recognize when arguments use evidence well
- **Causes** (.62)
  - Understand the difference between correlation and causation
  - Know when it is valid to infer that one thing causes another
- **Models** (.85)

- Use mathematical models to express ideas
- Manipulate mathematical models to draw conclusions
- Understand models and hypotheses that represent quantitative information

- **Quantitative** (.82)
  - Express ideas using quantitative information
  - Use quantitative information to solve problems
  - Use quantitative information to evaluate an argument

- **Statistics**[13]
  - Use data and statistics to evaluate factual claims

Table 3.3 gives for each of these dimensions estimates of the mean difference between students who did and did satisfy the QR–B requirement.[14] Positive figures indicate that QR–B students scored higher than their counterparts who did not satisfy the requirement, while negative values indicate they scored lower. The first column gives the unadjusted raw differences in mean response between the two groups of students; these estimates do not take account of the number of QR–B courses students took or academic performance in those courses. The remaining columns give the estimated effects of QR–B courses as computed from equation 1. Estimates were constructed for the "average" QR–B student, who took 1.4 courses, and for students who took only one QR–B course. Finally, the left-hand panel labelled "no controls" pertains to results generated without adjusting for other variables, while the right-hand panel displays the estimates obtained after adjusting for overall academic performance, sex, and QR–A exempt status. The rows of the table are arrayed in descending order of the magnitude of the adjusted estimate of the effect of QR–B on quantitative reasoning.

These results reveal clearly that the effect of fulfilling the QR–B requirement on quantitative reasoning is not of a piece. On the contrary, some skills that are conventionally marshaled under the

[13]Cronbach's reliability cannot be computed for a scale consisting of only one item.

[14]For continuity and comparability with the previous analyses, we focus on comparing the *mean scores* of QR–B students with the mean scores of other students. Yet because the individual items are measured on an ordinal scale, these types of dependent variables could be more properly analyzed within an ordered logit framework. Does the mode of analysis make a difference? As frequently happens, ordered logit models that correspond to the simple regressions fitted here yield results that are very similar both qualitatively and quantitatively to those reported below.

rubric of quantitative reasoning reveal large differences between those who did and did not satisfy the QR–B requirement, while other dimensions register no statistically detectable effect or even slightly negative effects.[15] The first two rows, which pertain to the scales Models and Quantitative, show statistically strong effects of QR–B, effects that get even stronger as we move from raw differences to differences that account for the number of QR–B courses students took and their performance in them. Compared to their counterparts who chose not to fulfill the QR–B requirement, students who satisfied the requirement report significantly higher levels of learning with respect to the use of models and quantitative information to express ideas, draw conclusions, and solve problems. These patterns are quite strong statistically, and buttressed by the findings in the third row. The statistical evidence in this row, though not as strong as rows 1 and 2, is by no means trivial, and suggests that QR–B students also gain more than other students in terms of learning how to use statistics and data to evaluate factual claims.

The scales labelled Models, Quantitative and Statistics largely exhaust the aspects of quantitative reasoning for which there is strong and reliable statistical evidence of a positive QR–B effect, i.e., an advantage in quantitative reasoning for those who satisfy the requirement. In contrast, the findings pertaining to the scales that reflect the acquisition of skills with respect to causal inference and the quality of arguments, provide little evidence of a systematic advantage for students who satisfied the QR–B requirement. On the contrary, as the last two rows of table 3.3 indicate, QR–B students tend to fall short of their comparable counterparts on both dimensions, although the pattern is statistically weak. Still, this evidence reinforces the findings on similar scales found in last year's QR–A study: students who fulfill the QR requirement report being at a slight disadvantage with respect to the "argument" and "causality" dimensions of quantitative reasoning. Still, the evidence of table 3.2 indicates that the gains achieved on the dimensions shown in the first three rows by students who

[15] The pattern of QR–B effects observed in this table closely parallels results found for QR–A in last year's study.

satisfy QR–B more than make up for the slight losses they report on the dimensions displayed in the last two rows.

One final issue that was explored concerned possible sex differences in the effects of satisfying the QR–B requirement. The conjecture is that the experience of satisfying the QR–B requirement may affect male and female students differently, so that they report gains of different magnitudes on one or more dimensions of quantitative reasoning. The summary results of the relevant statistical tests, carried out for each quantitative reasoning dimension, are displayed in Table 3.4. As these results convincingly demonstrate, there are no statistically detectable differences between male and female students in the effect of satisfying the QR–B requirement. Hence, the estimates displayed in table 3.3 apply, within sampling error, to male and female students alike.

It is important to remember that the findings of table 3.3 refer to *differences* between students who did and did not take a QR–B course in the way they assess the growth of their quantitative reasoning skills; the findings do not speak to the relative importance in the learning experience of each group of the different quantitative reasoning dimensions. Hence, it would be erroneous to infer that the observed advantage on the Models and Quantitative scales of those who satisfied the QR–B requirement means that those aspects of quantitative reasoning dominated the educational experiences of QR-B students during fall semester; similarly, it would be erroneous to infer that the slight advantage on the Arguments and Causes dimensions of those who did not satisfy the QR–B requirement means that those aspects of quantitative reasoning dominated the educational experiences of those students during fall semester. Table 3.5, which displays separately for each student group the absolute mean scores on the five scales, forcefully shows the problem with such conclusions. These mean scores indicate that, regardless of QR–B status, students rate their learning gains on the Models dimension lowest and their learning gains on the Arguments and Causes dimensions highest when evaluating what they learned as a result of their courses and other educational experiences during fall semester 2005. Both groups emphasize the skills acquired in *evaluating arguments* and in *making valid causal inferences*, although the quantitative and statistics dimensions are also emphasized by QR–B students. Among students who did not take QR–B, *the two highest rated items* are those, listed at the bottom of Table 3.4, that tap skill at evaluating arguments and knowledge of causal inference. Similarly, among students who did satisfy the QR–B requirement, *two of the three highest rated items* are those in the bottom rows of Table 3.5 that pertain to recognizing sound arguments and making valid causal inferences.

These findings are at once revealing and reassuring. They reveal that the ability to evaluate the soundness of arguments and to recognize valid causal inference are not skills unique to quantitative reasoning courses so much as they are part and parcel of reasoning itself, plain and simple. It is reassuring that, regardless of their QR–B status, students rate advances in their ability to reason as among the most important aspects of their UW learning experience. Another way to characterize students' highest rated experiences would be to say they report the largest improvements in their capacity for "critical thinking," since the Arguments and Causes scales capture the heart of that overused concept. As observed in a recent Wall Street Journal article on the subject[16],

> Critical thinking means being able to evaluate evidence, to tell fact from opinion, to see holes in an argument, to tell whether cause and effect has been established, and to spot illogic.

These are precisely the skills that the Arguments and Causes scales tap. We have a great deal of confidence in the general conclusion that all students tend to rate highest their improvement in critical thinking because it is corroborated by last year's study of QR–A, which yielded findings that very closely parallel those reported here. The results in Table 3.5 indicate that both groups of students, those who did and did not satisfy QR–B, credit their educational experiences during

[16]This observation was made by Sharon Begley in her Science Journal column, WSJ, October 20, 2006.

fall 2005 with generally contributing more to their critical thinking skills than to other aspects of quantitative reasoning.

### 3.3 Discussion

The findings above show that students who took a QR–B certified course during fall semester 2005-06 report greater improvement in their quantitative reasoning skills as a result of their UW educational experiences. They also show that the more QR–B courses a student took and the better their academic performance in them, the greater the reported gains in quantitative reasoning proficiency. The fact that average scores on the quantitative reasoning scales show an clear positive gradient with respect to numbers of courses and level of performance strongly suggests that these results reflect actual improvements as would have been detected by direct measures of learning outcomes.

Like the QR–A effects turned up last year, QR–B effects do not extend to all dimensions of quantitative reasoning broadly understood. On the contrary, the data clearly suggest that the quantitative reasoning skills that are improved by satisfying the QR–B requirement are those pertaining to the use and manipulation of models and quantitative information to express ideas. Such skills are probably the core of what is conventionally meant by "quantitative reasoning," so this is an important result. Nevertheless, other nontrivial and well-recognized, if less explicitly mathematical, dimensions of quantitative reasoning, including those pertaining to the principles of causal inference and the logic of argument, are largely untouched by QR–B effects. This overall pattern of QR–B effects closely follow those found in last year's studies of QR–A.

## 4. Conclusions

The results of this assessment study testifies to the effectiveness of the QR–B requirement as an instrument for achieving general education objectives. This study leaves no doubt that the growth in quantitative reasoning skills that students experience as a result of satisfying the QR–B requirement is appreciable and occurs in roughly equal measure for males and females. To be sure, the benefits of QR–B courses are limited to the more explicitly mathematical dimensions of quantitative reasoning. Even though the QR–B requirement may not yield dividends with respect to the non-mathematical aspects of quantitative reasoning, these other reasoning skills are definitely affected by the educational experiences of UW students. One important finding, corroborated by last year's QR–A study, is that all students, regardless of QR–B status, rate advances in their skill at evaluating the logic of arguments and understanding causal inference as among the most significant learning outcomes achieved at UW. These skills, which form the core of what is understood as "critical thinking," appear to be part and parcel of a UW education, and are available regardless of whether students choose a schedule of courses that satisfy the QR–B requirement.

**Table 2.1 Stratum Characteristics of the Samples Selected for the Study of the Effect of QR–B on Self-Reported Quantitative Reasoning Skills, Fall 2005-06**

| | *QR-B fall 2005?* | | |
|---|---|---|---|
| *QR-B prior to fall 2005?* | *No* | *Yes* | *Total* |
| *No* | 137 | 164 | 301 |
| *Yes* | 10 | 9 | 19 |
| *Total* | 147 | 173 | 320 |

**Table 2.2 Populations, Samples, and Respondents, by Stratum, for Study of the Effect of QR–B on Self-reported Quantitative Reasoning Skills, Fall 2005-06**

| | *Target Sample Size* | *Actual Sample Size* | *Respondent Sample Size* | *Response Rate* |
|---|---|---|---|---|
| *QR–B stratum* | 160 | 161 | 123 | 76.4% |
| *Non-QR–B stratum* | 160 | 124 | 91 | 73.4% |
| *Total* | 320 | 285 | 214 | 75.1% |

**Table 2.3 Rate of Response by Number of QR-B courses, Fall 2005-06**

| | *Number of QR-B Courses* | | | |
|---|---|---|---|---|
| *Respondent* | *0* | *1* | *2* | *3* |
| *No* | 26.6% (33) | 25% (27) | 18.4% (7) | 26.7% (4) |
| *Yes* | 73.4% (91) | 75% (81) | 81.6% (31) | 73.4% (11) |
| *Total* | 100% (124) | 100% (108) | 100% (38) | 100% (15) |

**Table 2.4 Rate of Response by Gender, Fall 2005-06**

| | *Gender* | |
|---|---|---|
| *Respondent* | *male* | *female* |
| *No* | 27.5% (30) | 23.3% (41) |
| *Yes* | 72.5% (79) | 76.7% (135) |
| *Total* | 100% (109) | 100% (176) |

**Table 3.1 The Effect of QR-B Courses on Students' Self-Reported Gains in Quantitative Reasoning (n=214)**

| *Independent variables* | *Models* 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| *QRB course (1=yes)* | .179* (1.77) | | | |
| *QR-B courses and grades* | | | | |
| *Number of QRB courses* | | .147*** (3.00) | .069 (1.18) | .086 (1.41) |
| *GPA in QR-B courses* | | | .079** (2.02) | .056 (1.30) |
| *Fall semester gpa* | | | | .151 (1.79) |
| *Constant* | 3.29 (41.24) | 3.28 (46.70) | 3.18 (34.71) | 2.74 (10.53) |
| *F-statistic* | 3.12 | 9.00 | 6.07 | 5.53 |
| *p-value* | .078 | .003 | .003 | .001 |

*Notes: The absolute t-ratios that appear below the estimated effects are based on robust standard errors. In model 4, the coefficients of number of QR-B courses and grade point average in QR-B courses are jointly significant, with F=4.6 and p-value=.01.*

** $p < .10$ **$p < .05$ ***$p < .01$*

**Table 3.2 Mean Quantitative Reasoning Scores by Number of QR-B Courses, fall 2005 (n=214).**

| *Number of QR-B courses* | *QR scale mean* |
|---|---|
| *0* | 3.29 (91) |
| *1* | 3.39 (81) |
| *2* | 3.58 (31) |
| *3* | 3.78 (11) |

Figure 1. Mean QR-B Scores by Number of QR-B Courses

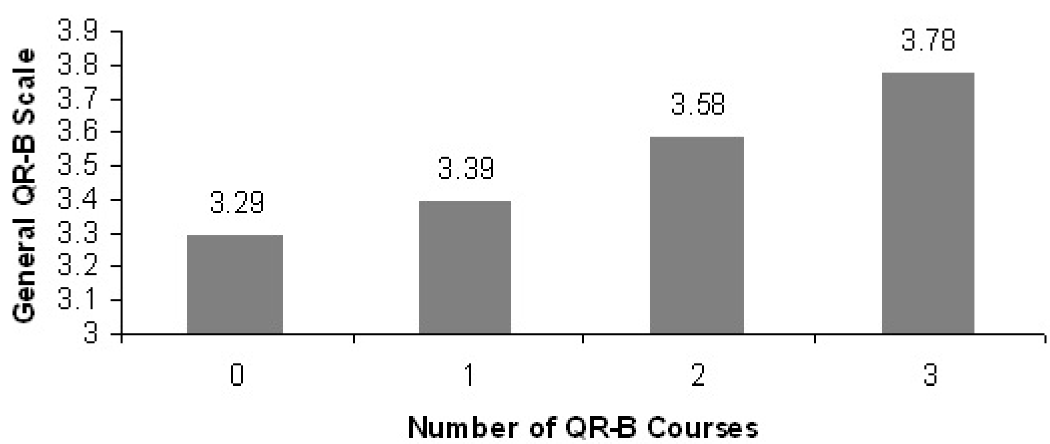

**Table 3.3 The Effect of QR-B Certified Courses on Students' Self-Reported Gains along Selected Dimensions of Quantitative Reasoning, Fall 2005-06 (n=214).**

| | | *Estimated Average Effect of QR-B on Dimensions of Quantitative Reasoning* | | | |
|---|---|---|---|---|---|
| | | *No controls* | | *Controls* | |
| *Quantitative reasoning dimensions* | *Raw Mean Differences (1)* | *All QR-B students (2)* | *One QR-B course (3)* | *All QR-B students (4)* | *One QR-B course (4)* |
| *Models* | *.587^^* *(4.12)* | .753^^ (5.89) | .621^^ (5.30) | .714^^ (5.19) | .583^^ (4.72) |
| *Quantitative* | *.272*** *(2.25)* | .394^^ (3.32) | .340^ (3.15) | .376^ (2.88) | .314^ (2.70) |
| *Statistics* | *.221* *(1.37)* | .342** (2.21) | .316** (2.32) | .265 (1.48) | .243 (1.54) |
| *Causes* | *.164* *(1.25)* | -.036 (0.30) | -.068 (0.05) | -.094 (0.69) | -.065 (0.53) |
| *Arguments* | *-.250*** *(2.28)* | -.106 (1.00) | -.068 (0.71) | -.170 (1.48) | -.134 (1.30) |

Notes: The absolute t-ratios that appear below the estimated effects are based on robust standard errors. "Raw differences" of column 1 are plain mean difference between the two groups. The other figures give estimates from equation *** in the text. The model without controls is equation 8888 in the text. The model with controls includes semester grade point average, QR-A exemption status, and sex.

* p < .10    **p < .05    ^p < .01    ^^p < .001

**Table 3.4 Tests of Sex Differences in Effects of QR-B Experience and Performance on Dimensions of Quantitative Reasoning Skill**

| *Quantitative reasoning dimensions* | *F-statistic* | *p-value* |
|---|---|---|
| *Models* | 0.39 | 0.68 |
| *Quantitative* | 0.73 | 0.48 |
| *Statistics* | 1.20 | 0.30 |
| *Causes* | 1.46 | 0.23 |
| *Arguments* | 1.28 | 0.28 |

**Table 3.5 Mean Ratings of Quantitative Reasoning Dimensions, by QR--B Status (n=214).**

| | **Satisfied QR-B Requirement?** | |
|---|---|---|
| ***Quantitative reasoning dimensions*** | **Yes** | **No** |
| ***Models*** | **3.31** | **2.72** |
| ***Quantitative*** | **3.51** | **3.23** |
| ***Statistics*** | **3.50** | **3.27** |
| ***Causes*** | **3.51** | **3.68** |
| ***Arguments*** | **3.62** | **3.87** |