IDENTIFICATION OF UNSUCCESSFUL STUDENTS IN GENERAL CHEMISTRY

G. Robert Shelton*, Joseph Simpson**, and Diana Mason***
*Chemistry Program, Texas A&M University – San Antonio, San Antonio, Texas, USA, 78224
**Department of Social Sciences, Texas A&M University – San Antonio, San Antonio, Texas, USA, 78224
***Department of Chemistry, University of North Texas, Denton, Texas, USA, 76203

Corresponding author e-mail: dmason@unt.edu

ABSTRACT

The Networking for Science Advancement (NSA) team collected data from multiple general chemistry courses at nine universities within a broad geographic setting in a majority-minority US state. Data include diagnostic scores on the Math-Up Skills Test (MUST), quantitative literacy/quantitative reasoning (QL/QR) quiz, along with student demographics, and overall course grades. From these data the team determined how automaticity skills in procedural arithmetic and quantitative literacy and reasoning can be used to predict success in lower-division chemistry courses. By expanding this dataset, we extended our investigations to discover what characterizes successful and unsuccessful students in general chemistry, first and second semesters (Chem I and II) categorizing by on- and off-sequence courses. Student characteristics studied include factors such as ethnicity, gender, location of residence, and employment status. In a short amount of required classroom time (approximately 35 minutes is needed for students to complete both assessments and a demographic survey), it is possible to identify students at the start of the semester who will struggle in general chemistry. The MUST is the preferred predictor but using the MUST and QL/QR together enhances predictability. *[African Journal of Chemical Education—AJCE 13(2), June 2023]*

INTRODUCTION

The Republic of Texas was established in 1836 and became a state of the United States of American (USA) on December 29, 1845, serving under two flags, the Republic of Texas flag and the USA flag until February 19, 1845, Statehood Day. Texas has always been concerned about the education of our students. In fact, the second president of the Republic of Texas (1838–1841) Mirabeau B. Lamar is called the Texas Father of Education. His most famous quote is a "Cultivated mind is the guardian genius of democracy" and can be found on The University of Texas' seal as the motto *Disciplina Praesidium Civitatis*. By Egyptian standards, about 5000 years older, Texas still has a lot to learn. Even though Texas is the second largest US state by land mass (only Alaska is larger), Egypt is 44% larger and has 79M more people. Texas is one of five majority-minority states in the USA with a minority population of 40% Hispanics, 13% Black, and about 7% other minorities leaving about 40% classified as White, non-Hispanic. The area of Texas covered in the studies by the Texas Networking for Science Team (NSA) can be seen in Fig. 1. The area covered by the NSA investigations is over 45,000 mi^2 or about 117,000 km^2. Within this area, six New England states (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont) could be placed with almost as much land in Texas still remaining.

Figure 1. Location of Texas institutions within the red-boxed area that participated in the NSA team studies.

Given the wide variety of participating institutions across a broad and diverse ethnic and geographic setting, the research population provides a representative view of an even larger population lending credibility to the study that reflects beyond what is typically reported for a single institution. In this study, we focused on evaluating students in general chemistry I and II (Chem I and II) who were unsuccessful (grades of D and F) and those deemed successful (grades of A, B, and C). The study compared how struggling students' automaticity skills or what they can do without a calculator differs from those of successful students. Two instruments were used to evaluate students' automaticity: the MUST (Math-Up Skills Test) and a QL/QR quiz that investigated their quantitative literacy/quantitative reasoning abilities. To broaden the applicability of this study, students enrolled in on- and off-sequence courses in Chem I or II were investigated. Typically, Chem

I on- and Chem II off-sequence courses are offered in the fall semester and Chem II on- and Chem I off-sequence courses are offered in the spring semester.

Initial Results

The MUST was inspired by a 16-question (16-Q) quiz in a publication by [1]. Since then, the NSA team has added four questions (Qs) to the original version stressing the arithmetic associated with using fractions. The MUST instrument has been used in multiple studies resulting in 13 publications and one more submitted manuscript [2-15] where it has been shown to give consistent and repeatable results. The MUST assesses basic overlearned procedural arithmetic skills of students when they not allowed to use their calculators for this 15 min., open-ended quiz the first week of class. A copy of the MUST can be found in [14]. Correctly solving the MUST exercises requires students to not only know the basic operations (add, subtract, multiple and divide) but also to know the procedures needed to correctly solve the problems. The QL/QR assessment does not require a calculator to solve the exercises. It assesses the ability of students to read and understand questions that require data usually in the form of images (graph, chart, diagram, etc.) to answer the 20 multiple-choice questions [11]. Students' QL/QR skills in our data-driven world are becoming a more and more important factors in students' education. Results show a strong correlation between students' automaticity MUST skills and their QL/QR abilities (r = 0.60) [11]. Published MUST results for predicting success of at-risk Chem I students is around 78% [14] and for Chem II students is about

83% [10]. Adding the QL/QR as an additional diagnostic quiz improved our ability to identify potentially about 9% more at-risk students [11].

METHODOLOGY/EXPERIMENTAL

Instruments

The MUST assesses a student's ability to conduct basic mathematical operations including multiplication, division, square roots, fractions, logarithms/ exponents, and symbol manipulation without the use of a calculator and has consistently produced strong reliability ($r_{\mathrm{KR20}} = 0.855$) and a very large effect size data (Cohen's $d > 1.2$). The KR-20 formula used to determine r follows, where k is the number of questions asked and p = percent correct, $q = 1 -$ percent correct, and σ^2 is the standard deviation squared: $r_{\mathrm{KR20}} = [k/(k - 1)][1 - (\Sigma pq/\sigma^2)]$.

The QL/QR quiz was specifically developed as an instrument where calculators would not be needed to answer the exercises. Many of the problems were selected from questions in Eric Gaze's database of questions (NSF DUE 1140562 project). The QL/QR assessment showed a medium effect size (Cohen's $d = 0.54$) and acceptable reliability of (KR-20 = 0.738). The exercises on the QL/QR consists of three distinct components: arithmetic, algebra, and problems with images. Analysis of these three components indicated that there is the existence of a very large effect size of problems with images on course averages ($d > 1.2$), but the overall effect size of the complete QL/QR is lower ($d > 1.6$) than that of the MUST on courses averages. When comparing the MUST and QL/QR scores, there is also a large effect size indicating that procedural arithmetic skills as measured by the MUST has a strong relationship to the skills needed to correctly solve the QL/QR exercises.

Participants

A population of n = 1,915 from nine institutions broken into subgroups based on their general chemistry enrollment status (on- and off-sequence) was evaluated: Chem I on (n = 735), Chem I off (n = 624), Chem II on (n = 381), and Chem II off (n = 175). The students attend public and private institutions, those located in small towns and metropolitan areas, and schools that are considered to be small (under 4,000 enrollees) to large (enrollment over 50,000). The lecture class enrollment ranged from around 30 to over 300 students. All students evaluated consented to participate in these IRB-approved studies. No constraints were dictated to any of the instructors at these schools; all were encouraged to teach the courses as deemed acceptable by their departments. Given the large number of students and the ethnic and geographic diversity, results are considered as more generalizable than results typically reported for a single institution.

See Table 1 for the demographic breakdown of this population students who did and did not succeed in Chem I and II. Table 1 is repeated (R) in terms of Table 1R to illustrate that students' course average, MUST and QL/QR means are aligned from high to low score averages. Also, in Table 1R, note that the percentage of unsuccessful students increased as their respective diagnostic scores decreased. Table 1R also points to the fact that the best students are those who are enrolled in Chem II on courses.

Table 1. Diagnostic course and quiz averages and numbers of unsuccessful students from these courses

Course	***n***	**Course Average (%) (*SD*)**	**MUST Mean (*SD*)**	**QL/QR Mean (*SD*)**	**Unsuccessful *n* (%)**
Chem I on	735	76.4 (15.9)	38.8 (24.3)	64.2 (17.0)	209 (28.4%)
Chem I off	624	69.0 (17.8)	34.6 (21.8)	59.6 (16.6)	263 (42.1%)
Chem II on	381	82.5 (12.6)	53.2 (24.8)	69.8 (17.0)	57 (15.0%)
Chem II off	175	64.3 (16.9)	30.1 (18.8)	59.4 (16.6)	116 (66.3%)
Overall	**1,915**	**74.1 (17.0)**	**39.5 (24.3)**	**63.4 (17.3)**	**645 (33.7%)**

Table 1R. Repeat of Table 1 to show align of course average and MUST and QL/QR means from high to low scores along with an increase of unsuccessful students as scores decrease

Course	***n***	**Course Average (%) (*SD*)**	**MUST Mean (*SD*)**	**QL/QR Mean (*SD*)**	**Unsuccessful *n* (%)**
Chem II on	381	82.5 (12.6)	53.2 (24.8)	69.8 (17.0)	57 (15.0%)
Chem I on	735	76.4 (15.9)	38.8 (24.3)	64.2 (17.0)	209 (28.4%)
Chem I off	624	69.0 (17.8)	34.6 (21.8)	59.6 (16.6)	263 (42.1%)
Chem II off	175	64.3 (16.9)	30.1 (18.8)	59.4 (16.6)	116 (66.3%)
Overall	**1,915**	**74.1 (17.0)**	**39.5 (24.3)**	**63.4 (17.3)**	**645 (33.7%)**

Of the 1,915 students, 645 students (33.7%) were not successful in their respective courses (Table 2). The score alignment found in the complete class (Table 1R) does not track to the subset of unsuccessful students where the trend no longer matches. All averages in Table 2 when compared to their corresponding entries in Table 1 are statistically lower ($p < 0.05$). Other demographic information gathered about students from a one-page, open-ended questionnaire includes whether or not students who lived on campus or not made a difference in their final course average and what impact did working have on students' course averages. Residence location did not make a difference but whether students did or did not work made a difference. The greatest negative effect on final course averages was due to working full time, but students who worked for only 10 h/week on campus had a slight positive grade boost. Females outperformed males in Chem I, but enrollees in Chem II on showed male students with higher course averages. For the most part, white non-Hispanics and Asians outperformed Hispanics and the other ethnicities.

Table 2. Diagnostic course and quiz averages for unsuccessful students

Course	***n***	**MUST Mean (*SD*)**	**QL/QR Mean (*SD*)**	**Course Average (%) (*SD*)**
Chem I on	209	26.8 (17.7)	56.4 (16.1)	56.7 (12.1)
Chem I off	263	27.6 (17.9)	55.2 (15.6)	52.7 (14.6)
Chem II on	57	32.2 (19.9)	62.0 (15.3)	61.1 (8.7)
Chem II off	116	26.3 (17.8)	56.4 (16.0)	55.4 (12.4)
Overall	**645**	**27.5 (28.0)**	**56.4 (15.9)**	**55.2 (13.2)**

RESULTS

Figs. 1 and 2 display charts for the MUST and QL/QR assessments, respectively. In all cases, the mean scores of each question on the MUST and QL/QR illustrate the same up and down patterns regardless of the class in which these unsuccessful students were enrolled (Chem I and II, on and off semesters). Considering that over 90% of these students attended a secondary school in Texas and were exposed to an isomorphic curriculum, it is noteworthy that they appear to hold similar misconceptions. In general, there is little observable difference between the diagnostic quiz's means of these unsuccessful students regardless of the course enrolled.

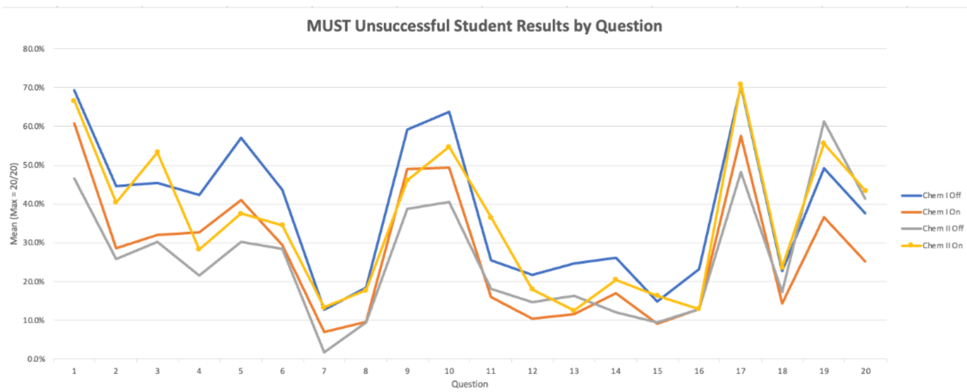

Figure 1. MUST exercises' means by question.

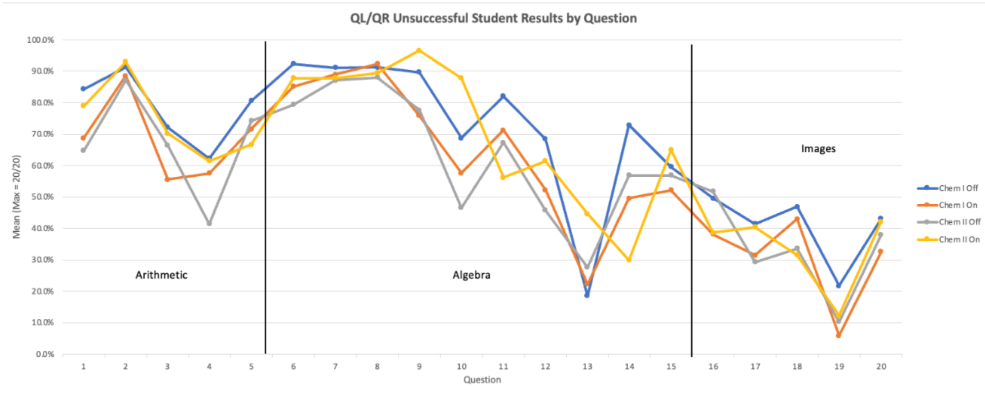

Figure 2. QL/QR exercises' means by question: questions (Qs) 1-6 assessed arithmetic, Qs 6-15 assessed algebra, and in Qs 16-20 used images (graphs, charts, diagrams, etc.) to solve the problems. In addition to the overall similar up and down pattern, there appears to be a downward trend of success from arithmetic exercises to problems that require the interpretation of images to be solved.

Another way to evaluate the data is to look the predictability of the MUST and QL/QR assessments using alluvial diagrams to display the results. The first task is to determine the middle score range for the MUST and QL/QR for each class. With the average known, subtract one-half the *SD* and add one-half the *SD* to that average. For example, if the average score is a 40% and the *SD* is 24, then the middle range is 40 – 12 = 28 and 40 + 12 = 52, resulting in a middle range of 28-52%; the under average range is for students who score below 28% and the above range is over 52%. See Table 3 for the categorial data (under, middle, above) for each course as to their MUST and QL/QR scores. Figs. 3-6 are the supporting alluvial diagrams for each course. Can students score in the above average range on the MUST and the QL/QR and still be unsuccessful in the course? Yes! Can students score under average on the MUST and the QL/QR and still be successful in the course? Yes! BUT the odds are against you. In Fig. 3 for Chem I on-sequence students, follow the blue river from the left side to the middle and note the much smaller percentage of students who scored above average on the MUST and were not successful. This flow is consistent in Figs. 4-6. The QL/QR does not produce as clear of picture until Fig. 6 where it is obvious that the students who were not successful were the ones who not only had under average MUST scores but also were the majority of the unsuccessful QL/QR students (note the purple and orange rivers).

Table 3. MUST and QL/QR score ranges for the alluvial diagrams

	MUST Ranges (%)			**QL/QR Ranges (%)**		
Course	**Under**	**Middle**	**Above**	**Under**	**Middle**	**Above**
Chem I On	< 26.7	26.7 – 51.0	> 51.0	< 55.7	55.7 – 72.7	> 72.7
Chem I Off	< 23.7	23.7 – 45.5	> 45.5	< 51.3	51.3 – 67.9	> 67.9
Chem II On	< 40.8	40.8 – 65.6	> 65.6	< 61.3	61.3 – 78.3	> 78.3
Chem II Off	< 20.7	20.7 – 39.5	> 39.5	< 51.1	51.1 – 67.7	> 67.7
Gen Chem	**< 27.4**	**27.4 – 51.7**	**> 51.7**	**< 54.8**	**54.8 – 72.0**	**> 72.0**

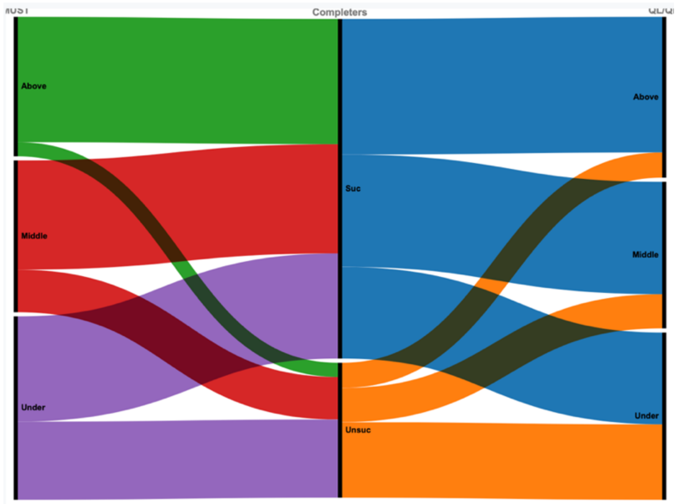

Figure 3**:** Alluvial diagram for Chem I on-sequence course. The MUST ranges are on the left-side bar and the QL/QR ranges are on the right-side bar. The center bar represents the blocks of students who were successful (Suc) and unsuccessful (Unsuc) in the course. Only a small percentage of these Unsuc students who entered with above average MUST scores were unsuccessful in the course (follow the green river from the left bar to the bottom of the center bar). A slightly greater percentage of the Suc students performed better on the QL/QR (blue river) than the MUST (green river). Over half of the Unsuc students scored under average on the MUST (purple river) and on the QL/QR (orange river). Source: https://www.rawgraphs.io/learning/how-to-make-an-alluvial-diagram

 ISSN 2227-5835

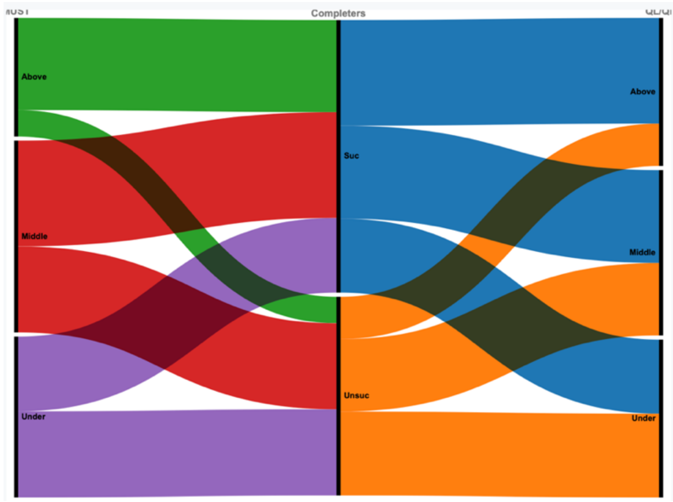

Figure 4: Alluvial diagram for Chem I off-sequence course. The MUST ranges are on the left-side bar and the QL/QR ranges are on the right-side bar. The center bar represents the blocks of students who were successful (Suc) and unsuccessful (Unsuc) in the course. For this group of students, the notable observation is that the students who were above on the MUST (green river) were more likely to succeed than not. Source: https://www.rawgraphs.io/learning/how-to-make-an-alluvial-diagram

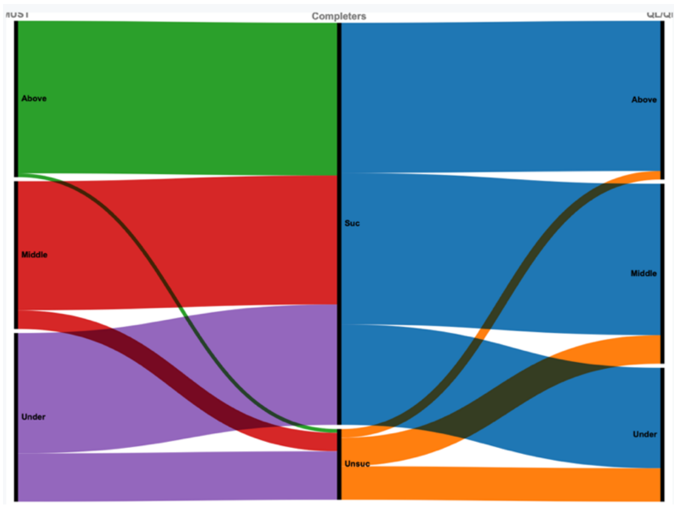

Figure 5: Alluvial diagram for Chem II on-sequence course. The MUST ranges are on the left-side bar and the QL/QR ranges are on the right-side bar. The center bar represents the blocks of students who were successful (Suc) and unsuccessful (Unsuc) in the course. Very few students who scored in the above average range on the MUST (green river) were Unsuc in the course and likewise with the students who performed well on the QL/QR (orange river). However, there was a significant percentage of students who scored under average on the MUST (purple river) and under average on the QL/QR (blue river) who succeeded in the course probably due to their improved background from successful completion of Chem I. Source: https://www.rawgraphs.io/learning/how-to-make-an-alluvial-diagram

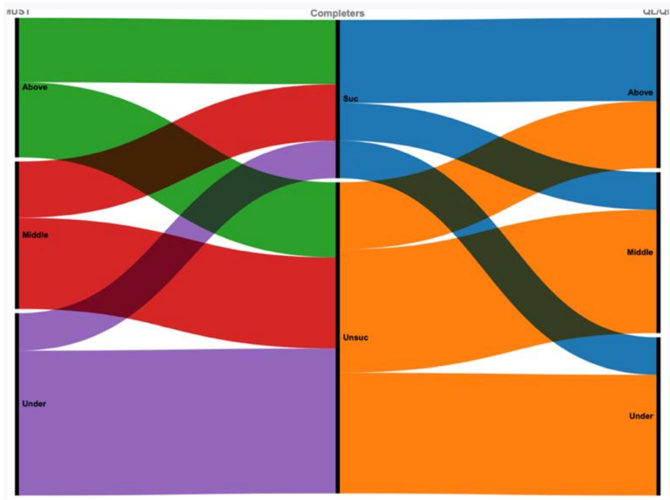

Figure 6: Alluvial diagram for Chem II off-sequence course. The MUST ranges are on the left-side bar and the QL/QR ranges are on the right-side bar. The center bar represents the blocks of students who were successful (Suc) and unsuccessful (Unsuc) in the course. There were more Unsuc students than Suc students in this course (center bar). About a quarter of the Unsuc students scored in the above average range on both the MUST (green river) and QL/QR (orange river). About half of the Unsuc students scored in the under average range on both diagnostics (purple and orange rivers). Source: https://www.rawgraphs.io/learning/how-to-make-an-alluvial-diagram

Research Question

To what extent are the data from the MUST and QL/QR diagnostic instruments statistically predictable of success in Chem I and Chem II, on- and off-sequence courses.

Data from this study were split into five pairs of unequal samples. The first sample consisted of a balanced random selection of Chem I students on and off sequence and Chem II students on and off sequence to ensure that both the training and validation samples contain balanced proportions from each student group. Students with missing data (one of the diagnostics not available) were deleted leaving three-fourths of a full sample as n = 1,303 used for the training model and the remaining one-fourth (n = 433) to be held out to test the accuracy of the model's prediction [16]. Table 4 list samples consisting of Chem I and II, on and off sequence divided into training and validation samples. The LASSO method is a regression analysis method that regularizes, smooths, and shrinks model covariates in an effort to find the set of model coefficients that optimize prediction accuracies in balance with predictive effects for subject covariate variables [16,17]. The linear model uses cross validation selection criteria to minimize the function's estimate of the mean square error (MSE). As a consequence, it selects the most parsimonious model with the largest out-of-sample explained variance. R^2 values between 0.3-0.5 are moderate correlations.

Table 4. Goodness of fit for linear and logistic LASSO regression predictive models

Model	*Sample*	*MSE*	*R^2*	*Observations*
LASSO Linear	Training	194.9540	0.3016	1,303
	Validation	220.6942	0.2943	433
MUST Only	Training	222.6510	0.2024	1,303
	Validation	244.0774	0.2195	433
QL/QR Only	Training	238.9954	0.1438	1,303
	Validation	273.5914	0.1251	433
Chem I On				
LASSO Linear	Training	168.0483	0.3398	500
	Validation	195.5938	0.2406	166
MUST Only				
	Training	193.5125	0.2398	500
	Validation	219.9153	0.1462	166
QL/QR Only				
	Training	211.3044	0.1699	500
	Validation	223.6718	0.1316	166
Chem I Off				
LASSO Linear	Training	256.5008	0.1620	431
	Validation	285.0799	0.1290	143
MUST Only				
	Training	273.8302	0.1054	431
	Validation	287.2640	0.1223	143
QL/QR Only				
	Training	284.628	0.0701	431
	Validation	306.5000	0.0636	143
Chem II On				
LASSO Linear	Training	111.7107	0.2708	258
	Validation	112.2044	0.2458	86
MUST Only				
	Training	113.1920	0.2611	258
	Validation	111.0082	0.2538	86
QL/QR Only				
	Training	137.3789	0.1032	258
	Validation	119.5837	0.1962	86
Chem II Off				
LASSO Linear	Training	262.0698	0.0498	114
	Validation	272.5452	0.0241	38
MUST Only				
	Training	253.1177	0.0823	114
	Validation	273.6034	0.0204	38
QL/QR Only				
	Training	266.9475	0.0322	114
	Validation	235.2931	0.1575	38

Derived from a post-selection model with un-penalized coefficients.

The process of finding the LASSO penalty parameter (*lambda*, λ) that minimizes MSE in linear regressions is visualized in Fig. 7. In the graph, the *y*-axis starts with the smallest MSE from a cross-validation function containing no coefficients. As the curve moves along the *x*-axis, the MSE is reduced as λ shrinks to the lowest penalty before the MSE increases. In Fig. 8 graph, the selection of λ corresponds directly to the number of covariates included in the predictive models and the strength of their coefficients. The MUST score is the first covariate selected and has the largest contribution to the prediction. The QL/QL (Fig. 8) however does not perform as well at scores lower than 30%, but it does become a better linear predictor above that.

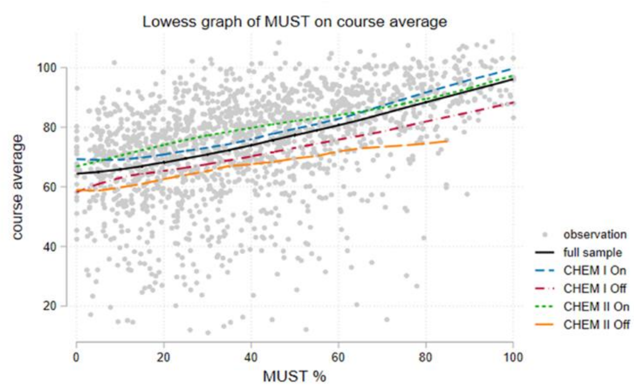

Figure 7. Course average vs. MUST percentage correct.

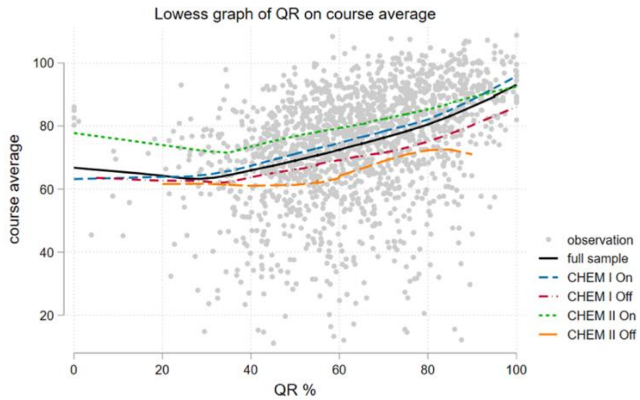

Figure 8. Course average vs. QL/QR percentage correct.

LASSO regression is not normally used for inference, but it is possible to select certain variables of interest to estimate the standard errors for the inputs. In this case the MUST and QL/QR scores are the variables of interest. Cross-fit partialing-out functions by splitting the sample and using one sample to calculate the LASSO linear regression coefficients in the second. To avoid bias several samples, in this case 10, are drawn and the results are averaged [17,18]. Table 5 presents the results from the full sample as well as the Chemistry I & II on- and off-sequence subsamples. MUST and QL/QR coefficients are directly comparable. In each sample, the MUST outperforms QR as a predictor, but the QL/QR still contributes to predictability of the final course averages.

Table 5. Cross-fit partialing-out LASSO linear regression coefficients

	Full Sample	Chem I On	Chem I Off	Chem II On	Chem II Off
MUST %	0.195***	0.217***	0.209***	0.189***	0.178*
	(0.0199)	(0.0283)	(0.0405)	(0.0295)	(0.0755)
QL/QR %	0.158***	0.196***	0.112*	0.107**	0.160
	(0.0260)	(0.0412)	(0.0551)	(0.0343)	(0.0960)
Observations	1736	666	574	344	152

Robust standard errors are in parentheses; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Fig. 9 use a lowess smoother to visualize the difference in explanatory power by removing the noise and creating a smooth line to help visualize the relationship between the variables influence on the course average. The MUST is a better predictor of course average having a consistently linear relationship across observations. This indicates that on average a student who performs poorly on the MUST will tend to have a lower course average and those who perform well will have higher course averages.

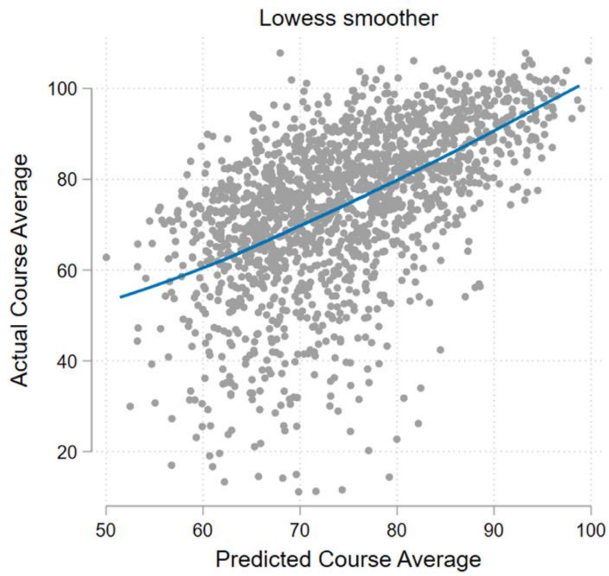

Figure 9**.** Actual vs. predicted course average showing a positive slope.

RESULTS AND DISCUSSION

Limitations

While the full sample is the most reliable model due to its sample size, most of the subsamples also work well in this case, except the Chem II off-sequence subsample whose estimates are likely not reliable (see Chem II off-sequence curved line in Fig. 8). This observation is consistent with Chem II off-sequence subsample being the lowest performing group overall and the group with the largest percentage of unsuccessful students (see Fig. 6). Reasons for students not succeeding are many from lack of academic preparation to emotional family situations. Without personal interviews this inquiry is not possible.

Conclusions

Rapid technological and social changes are creating a more interconnected world that is growing more diverse. We are preparing general chemistry students for global competence. Students' dependence on technology is hurting their quantitative literacy and reasoning abilities. Digital natives cannot make up for a lifetime of using technology, but can be provided opportunities in the classroom to solve some exercises without the calculator so that skills of estimating answers can be practiced.

Can you identify general chemistry students at the start of the semester who will struggle with the course? YES! If you can only give one diagnostic, the MUST is the better of the two diagnostic instruments (Fig. 10). Giving both MUST and QL/QR improves the chances of identifying about 10% more students who are at-risk of not succeeding in general chemistry. The more emphasis that is placed on QL/QR the better students will be prepared for this data-driven world. Chem II on-sequence students appear to be the best prepared to succeed. Using these students as the model, the more students' mental-math skills are honed, the more successful all students will be. Of the prepared students, 88.3% of Chem I on-sequence students and 90.5% of Chem II on-sequence students were successful. In this study, we drew inferences between procedural arithmetic and QL/QR skills from the results of two diagnostic instruments (Fig 10). Fig. 10 uses a concept map to illustrate how the diagnostic assessments' statistical values from Table 5 support the strong relationship between low scores on the assessments and failing to be successful in the course and vice versa. Using the MUST and the QL/QR diagnostics, about half of the students who are unsuccessful in Chem I and II present early warning signals that can be uncovered in a minimal amount of class time at the beginning of a semester.

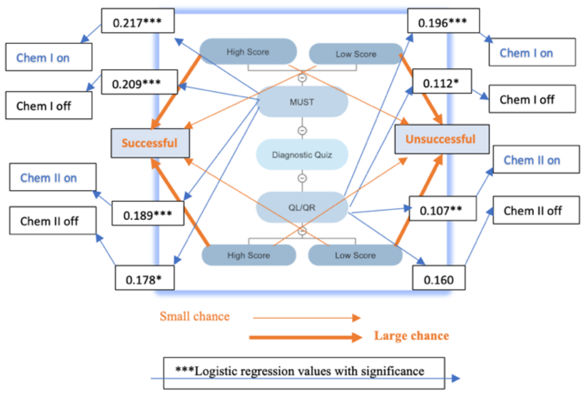

Figure 10. Concept map of LASSO linear regression coefficients on Chem 1 and II, on and off semesters.

REFERENCES

1. Hartman, JA. R.; Nelson, E. A. (2016). Automaticity in computation and student success in introductory physical science courses. Cornell University Library. arXiv:1608.05006v2 [physics.ed-ph] Paper presented as part of Chemistry & Cognition: Support for Cognitive-Based First-Year Chemistry, 2016 (accessed May 14, 2022). Link to quiz: http://bit.ly/1HyamPc
2. Albaladejo, J. DP.; Broadway, S.; Mamiya, B.; Petros, A.; Powell, C. B.; Shelton, G. R.; Walker, D. R.; Weber, R.; Williamson, V. M.; Mason, D. ConfChem Conference on Mathematics in Undergraduate Chemistry Instruction: MUST-know pilot study—math preparation study from Texas. *Journal of*

Chemical Education 2018, *95*(8), 1428-1429. (doi: 10.1021/acs.jchemed.8b00096) [Articles ASAP (As Soon As Publishable): July 20, 2018 (Report).]
3. Alivio, T. E. G.; Howard, E. H.; Mamiya, B.; Williamson, V. M. (2020). How does a math review impact a student's arithmetic skills and performance in first-semester general chemistry? *Journal of Science Education and Technology*, *29*(6), 703-712. https://doi.org/10.1007/s10956-020-09851-7
4. Bodenstedt, K.; Dubrovskiy, A.; Lee, K.; Rix, B.; Mason, D. (2022). Impact of students' automaticity ability on their success in o-chem I. *Biomedical Journal of Scientific & Technical Research*, *42*(1). (doi: 10.26717/BJSTR.2022.42.006700) https://biomedres.us/pdfs/BJSTR.MS.ID.006700.pdf
5. Dubrovskiy, A.; Broadway, S.; Jang, B.; Mamiya, B.; Powell, C. B.; Shelton, G. R.; Walker, D. R.; Weber, R.; Williamson, V.; Villalta-Cerdas, A.; Mason, D. (2022). Is the gender gap closing? *Journal of Research in Science Mathematics and Technology Education*, *5*(1), 37-57 doi: https://doi.org/10.31756/jrsmte.512
6. Ford, R.; Broadway, S.; Mason, D. (submitted). e-Homework and motivation for students' success in first-semester general chemistry. *Journal of Science Education and Technology*.
7. Lee, K. S.; Rix, B.; Spivey, M. Z. (2023). Predictions of success in organic chemistry based on a mathematics skills test and academic achievement. *Chemistry Education Research and Practice. DOI: 10.1039/D2RP00140C*
8. Mamiya, B.; Powell, C. B.; Shelton, G. R.; Dubrovskiy, A.; Villalta-Cerdas, A.; Broadway, S.; Weber, R.; Mason, D. (2022). Influence of environmental factors on success of at-risk Hispanic students in first-semester general chemistry. *Journal of College Science Teaching*, *51*(4), 46-57.
9. Petros, A.; Weber, R.; Broadway, S.; Ford, R.; Powell, C.; Hunter, K.; Williamson, V.; Walker, D.; Mamiya, B.; Del Pilar, J.; Shelton, G. R.; Mason, D. MUST-know pilot—math preparation study from Texas. ACS DivCHED CCCE (Committee on Computers in Chemical Education) online conference organized by Cary Kilner and Eric Nelson: https://confchem.ccce.divched.org/content/2017fallconfchemp2 (last accessed October 24, 2022). Week 1B: October 23–October 29, 2017.
10. Powell, C. M.; Simpson, J.; Williamson, V. M.; Dubrovskiy, A.; Walker, D. R.; Jang, B.; Shelton, G. R.; Mason, D. (2020). Impact of arithmetic automaticity on students' success in second-semester general chemistry. *Chemistry Education Research and Practice*, *21*, 1028-1041. doi: 10.1039/D0RP00006J
11. Shelton, G. R.; Mamiya, B.; Walker, D. R.; Weber, R.; Powell, C. A.; Villalta-Cerdas, A.; Dubrovskiy, A. V.; Jang, B.; Mason, D. (2021). Early warning signals from automaticity diagnostic instruments for first- and second-semester general chemistry, *Journal of Chemical Education*, *98*, 3061-3072. doi: 10.1021/acs.jchemed.1c00714
12. Villalta-Cerdas, A.; Dubrovskiy, A.; Mamiya, B.; Walker, D. R.; Powell, C. B.; Broadway, S.; Weber, R.; Shelton, G. R.; Mason, D. (2022). Personal characteristics influencing college readiness of Hispanic students in a STEM gateway course: first-semester general chemistry. *Journal of College Science Teaching*, *51*(5), 31-41.
13. Weber, R.; Powell, C. B.; Williamson, V.; Mamiya, B.; Walker, D. R.; Dubrovskiy, A.; Shelton, G. R.; Villalta-Cerdas, A.; Jang, B.; Broadway, S.; Mason, D. (2020). Relationship between academic

preparation in general chemistry and potential careers. *Biomed Journal of Scientific & Technical Research*, *32 (5), 25311-25323, DOI: 10.26717/BJSTR.2020.32.005312*

14. Williamson, V. W.; Walker, D. R.; Chuu, E.; Broadway, S.; Mamiya, B.; Powell, C. M.; Shelton, G. R.; Weber, R.; Dabney, A. R.; Mason, D. (2020). Impact of basic arithmetic skills on success in first-semester general chemistry. *Chemistry Education Research and Practice*, *21*, 51-61 DOI: 10.1039/C9RP00077A
15. Willis, W. K.; Williamson, V. M.; Chuu, E.; Dabney, A. R. (2021). The relationship between a student's success in first-semester general chemistry and their mathematics fluency, profile, and performance on common questions. *Journal of Science Education and Technology, 31*, 1-15. https://doi.org/10.1007/s10956-021-09927-y
16. James G.; Witten D.; Hastie T.; Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*, Springer, New York.Hastie, Tibshirani, & Wainwright.
17. Belloni, A.; Chen, D.; Chernozhukov, V.; Hansen, C. B. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica 80*, 2369–2429. https://doi.org/10.3982/ECTA9626.
18. Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, *21*, C1-C68. https://doi.org/10.1111/ectj.12097.

ACKNOWLEDGEMENTS

The authors thank all CER instructors for their participation in the Networking for Science Advancement team and their students who agreed to participate in these studies. Thank yous are also extended to the organizers of ACRICE 5 2022 for the invitation to share our data and results from the on-going research in Texas.