7-10-2023

Predictability of the Must (Math-Up Skills Test)

Diana Mason
University of North Texas

G. Robert Shelton
Texas A&M University-San Antonio, bob.shelton@tamusa.edu

PREDICTABILITY OF THE MUST (MATH-UP SKILLS TEST)

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

Corresponding author e-mail: dmason@unt.edu

ABSTRACT

In the USA for the most part, completion of a first-semester general chemistry (Chem I) course lays the foundation deemed necessary for understanding second-semester general chemistry (Chem II) topics. Successful completion of Chem I and II gives students permission to progress to organic chemistry I (O-Chem). A series of studies undertaken by the NSA (Networking for Science Advancement) Texas team began in 2016. Texas is one of five majority-minority states in the USA and hosts a significant Hispanic population. The purpose of this research line is to evaluate the influence of basic arithmetic automaticity (what students can do without a calculator) skills needed to succeed in lower-level chemistry. Over 9,000 students from nine universities have contributed to this research. Results suggest a strong correlation between procedural arithmetic preparation, automaticity, and student performance in Chem I, II, and O-Chem courses. The NSA collaborative uses the Math-Up Skills Test (MUST) as an assessment instrument along with student demographics to identify at-risk students from these contributing populations at the beginning of a course with high reliability (*KR-20* = 0.863) and effect size (Cohen's $d \geq 1.20$). The hand-graded MUST requires only 15 minutes of class time to administer and combined with specific demographic categories consistently predicts students' success rate in lower-level chemistry about 80 percent of the time therefore providing adequate time to identify and help at-risk students. This paper is about the evolution of the MUST and how following the NSA team's research line has advanced its use and interpretation. *[African Journal of Chemical Education—AJCE 13(2), June 2023]*

INTRODUCTION

One of the hallmarks for chemical education research (CER) is that the researchers choose a meaningful (i.e., significant, important) problem. When searching for a problem that others would find meaningful, sometimes you only have to observe what is around you. The Scholastic Aptitude Test (SAT) is an accepted college entrance exam. The problem observed in Texas was that the state's SAT scores (blue line, Fig. 1) were declining rapidly as compared to the mean scores of the USA (red line, Fig. 1). Wanting to investigate what the issue(s) might be and having read a 2016 article by Hartman and Nelson [1], it seemed reasonable that students' lack of automaticity (what they could do without a calculator) skills might be the source of the problem. Hartman and Nelson's CER had compared what students could do without a calculator to what they could do with the use of a calculator. Repeating this study might be interesting, so the NSA (Networking for Science Advancement) Texas team was formed expanding the studied population to more than a single university. Hartman and Nelson did not name their 16-question (16-Q) quiz, but the NSA team did. This 15-min., hand-graded, 20-Q instrument was named the MUST (Math-Up Skills Test). To date, the project has produced 13 published research papers with one more submitted [2-15] and in addition, the NSA team collectively has presented 44 oral presentations. Currently, Macmillan Publishers (Austin, Texas) is piloting an online version of the MUST through the online *Achieve* program of the Macmillan Learning System (see: https://www.macmillanlearning.com/college/us/digital/achieve).

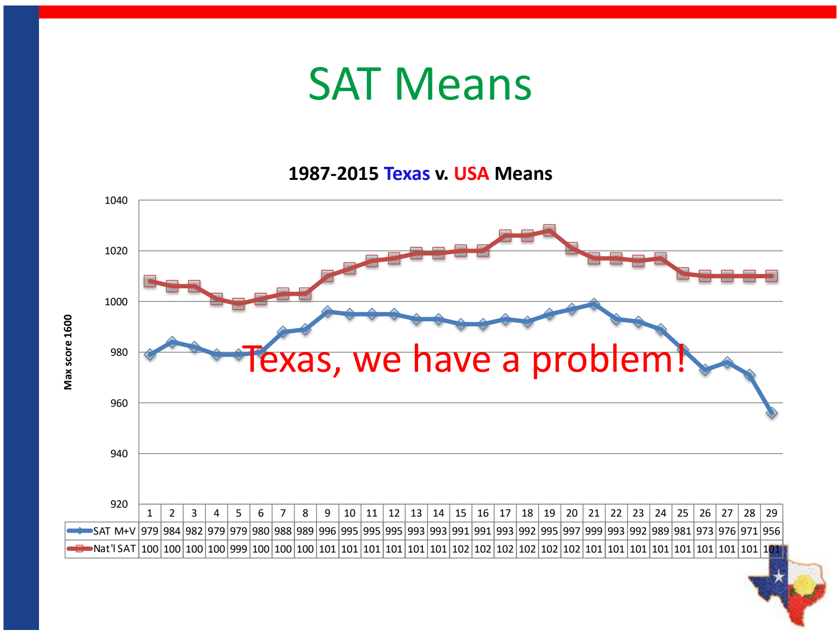


Figure 1. Comparison of Texas SAT annual means (blue) and SAT annual means in the USA (Petros et al., 2017).

The original team members were from six universities (blue line, Fig. 2) across the state of Texas (black border, Fig. 2). For comparison, Egypt (red border, Fig. 2) is about 1.5 times larger than the state of Texas. Both Austin, Texas and Cairo, Egypt are capital cities established on the same latitude of 30º north. The original team was composed of eight CER instructors, all with IRB

(Institutional Review Board) permission to conduct human subjects research from their respective institutions. The protocol was that general chemistry students would take the MUST twice, once without a calculator and then take a similar version with a calculator along with answer some general demographic questions. Attempting to discover what students could and could not do using a calculator was not only of interest to the authors but also to others who were concerned about the noted downward trend in the Texas SAT scores compared to those of the USA (Fig. 1). Another known fact was that calculator usage in Texas started as early as the seventh grade (middle school), so maybe students' automaticity skills not fully developed were being hampered by encouraging calculator use so early in the approved state curriculum.

Figure 2**.** Comparison of the size of Texas compared to the size of Egypt (1.5 times larger) and the territory (blue-dotted outline) covered by the NSA team from six institutions spread over 45,000 mi^2 or about 117,000 km^2 of the state of Texas, USA.

EXPERIMENTAL

Instrument

The 16-Q quiz [1] evolved to the 20-Q MUST (Math-Up Skills Test) after the pilot study [9] when the NSA team suggested that a few additional questions regarding the use of fractions needed to be added to the original quiz (for a copy of the 20-Q MUST see reference [14]). Over the past seven years, the diagnostic value of the MUST has produced some very interesting results for the NSA team. Three of the first data analyses that grab the team's attention are presented (see Figs. 3-5). The first eye-opener was that there was a stronger correlation between students' MUST scores and their final course averages when calculators were not used than when they were used (Figs. 3 and 5). Yes, students scored higher when they used calculators vs. when they did not (red vs. blue bars, Fig. 5), but the correlation to their final course average was stronger when they did not use a calculator (Fig. 3). The next observation that caught the team's attention was how the same "up and down pattern" of question means at each university stayed consistent (Fig. 4). These calculator-free, open-ended, hand-graded quizzes revealed that across the state students who had experienced an isomorphic curriculum (aka Texas Essential Knowledge and Skills) held similar misconceptions and had remembered (or not remembered) how to correctly solve certain arithmetic exercises. Yes, students from the premier post-secondary institutions performed at a higher level but the trendlines between all six institutions are very similar (Fig. 4).

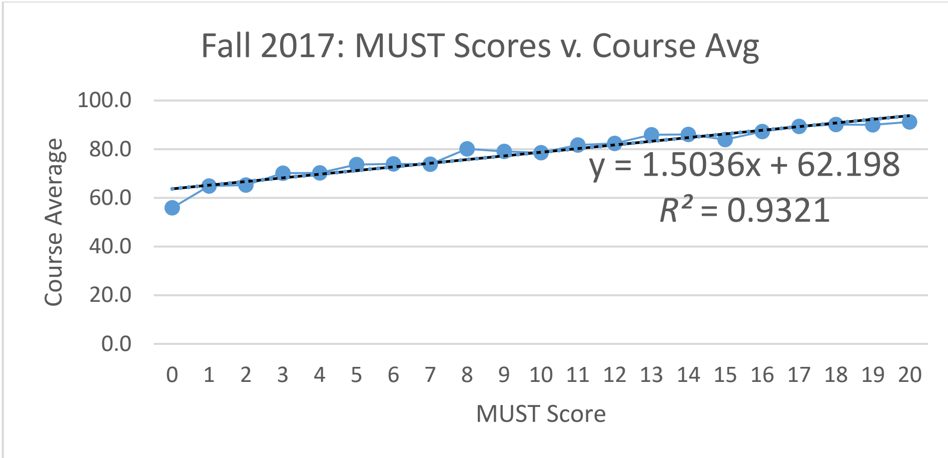


Figure 3. Course averages (n = 1,415) and their relationship to MUST scores without using a calculator.

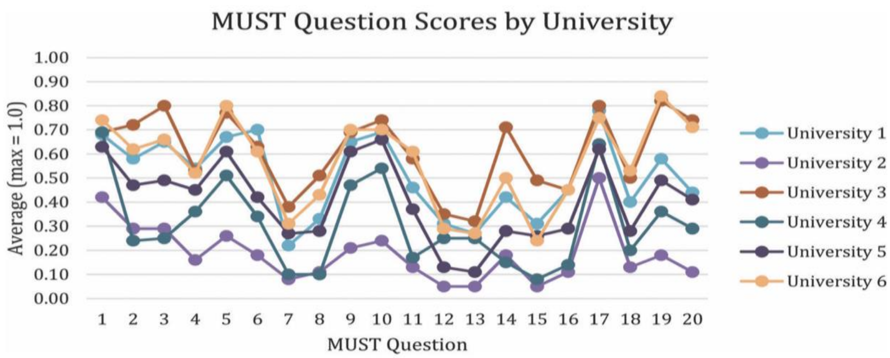
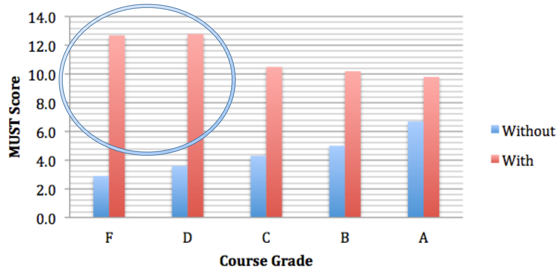


Figure 4. Mean scores for each question on the MUST. Note the similar up and down trends between each question mean at the six participating institutions (n = 1,073).



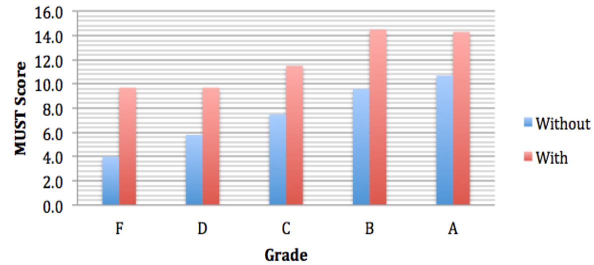


Figure 5**.** Bar graph of Chem I (left) and Chem II (right) of students' course grades compared to their average MUST scores. The most interesting trend (circled) was the Chem I students with a calculator (red bars): Students who were unsuccessful in the course (grades of D, F) scored higher than the successful students (grades of A, B, C).

The initial reaction was that this quiz simply covered basic arithmetic so this must be the problem behind students' lack of success in general chemistry. Some of the weakest students in Chem I can perform very well on the MUST (circled red bars, Fig. 5 left side), but they have very weak automaticity skills or what they can do without a calculator (blue bars, Fig. 5 left side). These students must be compensating for their lack of basic arithmetic skills by undoubtedly depending on the calculator. Since 2016, the MUST has evolved, the noted trends and correlations have remained the same, but our insights due to the NSA team's continued research have broadened.

METHODOLOGY

The first week of each semester students responded to 20-Q, timed MUST (15 minutes) and, during additional time (10 min.), complete a demographics survey. There are two versions of the MUST that differ only in minor number changes. For example, Q1 is 87 × 96 on one version and 78 × 96 on the other. The two versions of the MUST were validated by two mathematics professors, one chemistry professor, one chemistry education professor, and one science educator substantiating that it measured what it was intended to, and that it was appropriate for students enrolled in general chemistry. To date, no statistical difference (*t*-test) between versions has been observed.

Participants

Each instructor at the various institutions emailed deidentified student data to the NSA team lead, the first author of this paper. All unusable student data were eliminated (e.g., post-baccalaureates (small group), students with final course grades of < 10% (lack of participation), any student without a score for the MUST diagnostic quiz (absence), and students who did not give IRB consent to participate). Student demographics included that about 95% attended Texas secondary schools, 60% were freshman, 40% male, 30% Hispanic ethnicity, and about 50% worked either part- or full-time. Only 1% of the students did not take a high school chemistry course; however, almost 20% of the general chemistry students failed to meet the stated (suggested) mathematics prerequisites for enrollment.

Three more institutions have joined the NSA team bringing the number of members to 16. Over the past seven years of research studies, over 10,000 students have consented to participate. Of these, n = 9,315 Chem I and II students satisfied the research criteria mentioned above: Chem I = 6,303 (67.7%) and Chem II = 3,012 (32.3%). All studies until 2022 evaluated only Chem I and II students. In 2022, the investigations were expanded to include organic chemistry I (O-Chem I) as a separate population.

RESULTS

After choosing a meaningful problem and getting some interesting preliminary results, the next hallmarks to be met are: (1) Are the data statistically reliable and can the results be repeated? and (2) To what extent does the MUST predict which students will have a satisfactory course average (69.5% or higher)? Table 1 is a list of some of the publications by the NSA team members and the reliability and predictability statistics. Different studies emphasized different research questions, so not all of the same statistical data were generated for each study justifying why some of the data is missing in Table 1.

Table 1. Chem I and II statistical data: effect size, reliability, and predictability of the MUST

Publication	Date	Subject of Publication	Pop.	Effect size (Cohen's *d*)	Reliability (internal consistency)	Predictability
Petros et al. [9]	2017	Math preparation	2,127		KR_{21} = 0.821	
Albaladejo et al. [2]	2018	Math preparation	2,127		KR_{21} = 0.821	
Williamson et al. [14]	2020	Chem I student success	1,073		Cronbach's alpha = 0.856	78%
Powell et al. [10]	2020	Chem II student success	1,599	0.962	Cronbach's alpha = 0.853	83%
Weber et al. [13]	2020	Careers	4,113	1.43 Chem I; 1.20 Chem II	KR_{20} = 0.874	
Alivio et al. [3]	2020	Chem I impact of math review	325			
Shelton et al. [11]	2021	Chem I and II warning signals	1,915	1.21	KR_{20} = 0.855	
Dubrovskiy et al. [5]	2021	Gender gap	6,694	1.43 Chem I; 1.20 Chem II	KR_{20} = 0.874	
Villalta-Cerdas et al. [12]	2022	Personal characteristics of unsuccessful Hispanics	69	1.40	KR_{21} = 0.856	
Mamiya et al. [8]	2022	Environmental characteristics of unsuccessful Hispanics	69	1.40	KR_{21} = 0.856	80%
Willis et al. [15]	2022	Chem I common questions; linear and logistic regression models	1,020		Cronbach's alpha = 0.85	83.4%
Lee, Rix, & Spivey [7]	2022	Organic Chemistry	123	1.29	Cronbach's alpha = 0.861	82%
Ford, Broadway, & Mason [6]	submitted	Chem I e-homework	273	1.22	KR_{20} = 0.845	44%

Note: KR_{20} and KR_{21} (Kuder and Richardson 21) is a simplified version of *KR-20* that can be used when the difficulty of all items on the test are known to be equal. After analyzing data for multiple studies, the *KR-20* is a better choice for determining the reliability of these data that analyzed a binary or dichotomous choice (right/wrong) score on the hand-graded MUST. A high KR value indicates a stronger relationship between items as to their inter-item consistency. Like Cronbach's Alpha, 0.70 and above is good, 0.80 and above is better and 0.90 and above is the best, but above 0.90 also suggests that some items are redundant and make the data analyzed questionable.

Table 2 documents the most dramatic results for the population (n = 9,315): Chem I (top) and Chem II (bottom) students who were unsuccessful in the courses (grades of D or F) had limited automaticity skills based on MUST scores (maximum score = 20) and performed significantly lower than those who were successful ($p < 0.05$). Both successful and unsuccessful Chem II students did perform slightly better than successful and unsuccessful Chem I students, but still even the successful Chem II students averaged 11.40/20 = 57% on the MUST without a calculator.

Table 2. Performance on the MUST for successful and unsuccessful students

Chem I (*n* = 6,303)	***n* (Course Avg.)**	**MUST Score (*SD*)[a]**
Successful (≥ 69.5%)	4,356 (69.1%)	9.26 (4.95)
Unsuccessful (< 69.5%)	1,947 (30.9%)	5.70 (4.16)
Chem II (*n* = 3,012)	***n* (Course Avg.)**	**MUST Score (*SD*)[a]**
Successful (≥ 69.5%)	2,134 (70.8%)	11.40 (4.43)
Unsuccessful (< 69.5%)	878 (29.2%)	7.73 (4.67)

[a] Successful students performed significantly higher than unsuccessful students ($p < 0.05$).

Research Question #1

To what extent are the data from the MUST scores statistically reliable and can the results be repeated?

As can be seen in Table 1 above, the reliability data (based on Cronbach's alpha and KR data) and effect size data (Cohen's *d*) have consistently produced repeatable values. Alluvial diagrams (i.e., rivers showing *associations between categorical variables*) are constructed from the following online resource: https://app.rawgraphs.io. Fig. 6 is one of many alluvial diagrams that have been generated from NSA team data. The average MUST range was determined from the mean score and one standard deviation (*SD*) around the mean. The average range was determined by taking one half of the *SD* on either side of the mean. Possible scores on the MUST range from 0-20. For example, if the mean was 6.0 and the *SD* 4.0, then the average range is between 6 – 2 and 6 + 2 or a range of 4-8 leading to the low MUST range of 0-3 (L = low) and an above average range (U = upper) of 9-20. Can students in the upper MUST range make a D or F in the course? Yes! Can students in the lower level on the MUST succeed in the course? Yes! But the odds are that if you have skills that allow you to correctly respond to basic arithmetic problems without a calculator, you will succeed (follow the top violet river, Fig. 6); if students perform low on the MUST (follow the salmon-colored river starting on the bottom left, Fig. 6), a few do succeed but over half of the D's and F's flow from this group.

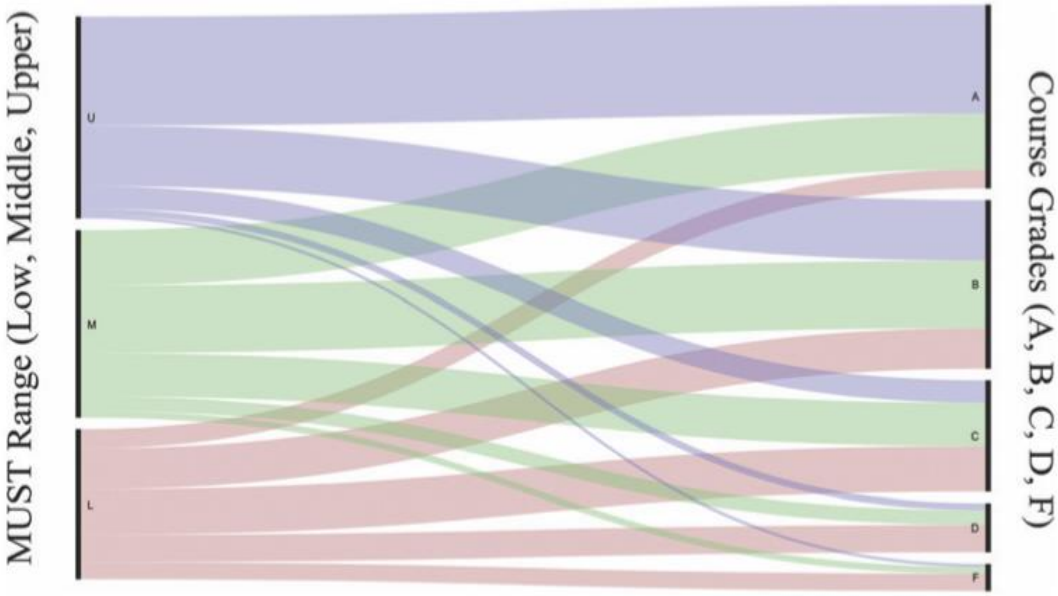


Figure 6. Chem II student data comparing MUST ranges of upper (U), middle (M), and lower (L) scores linked to students' final course grade. Follow the "rivers" to explore how each range of scorer performed.

Table 3 supports how the MUST scores correspond to final course grades of students in general chemistry. Must scores can also be used to predict students' success or failure in the respective classes. For more information regarding our predictability LASSO (Least Absolute Shrinkage and Selections Operator) models see the published results for Chem I [14] and Chem II [10].

Table 3. Classic averages and corresponding MUST scores for Chem I and II combined

Grade average	*n* (%)	MUST (*SD*)[a]
A: 89.5-100.0+%	1,985 (21.3%)	12.22 (4.47)
B: 79.5-89.4%	2,352 (25.2%)	9.95 (4.63)
C: 69.5-79.4%	2,154 (23.1%)	7.87 (4.59)
D: 59.5-69.4%	1,312 (14.1%)	6.62 (4.48)
F: 0-59.4%	1,512 (16.2%)	6.05 (4.36)
Overall (76.0%)	**9,315**	**8.88 (5.04)**

[a] Statistical difference between all nearest grade groups (A to B, B to C, etc.)

O-Chem Results

Having spent the first six years of research investigating Chem I and II students' automaticity skills, the conclusion was that the MUST results supported that the problem is in students' basic arithmetic preparation and their lack of automaticity skills. One of the NSA team members relayed the result that students who could do basic arithmetic problems without a calculator were by far the best students on to her organic chemistry (O-Chem) colleagues. Out of curiosity, the O-Chem professors gave the MUST to their students and the same results were obtained [7]: the students who were better on the MUST did better in O-Chem! Not fully convinced, the study was repeated with three universities pooling results [4]. In Fig. 7, the MUST ranges I (above average MUST) – III (below average MUST) (left side) were linked to whether students were S (successful) or U (unsuccessful) in O-Chem I. The results were even more defined than for Chem I and II—very few students in MUST group I were unsuccessful in O-Chem I. Why did the unexpected results produced?

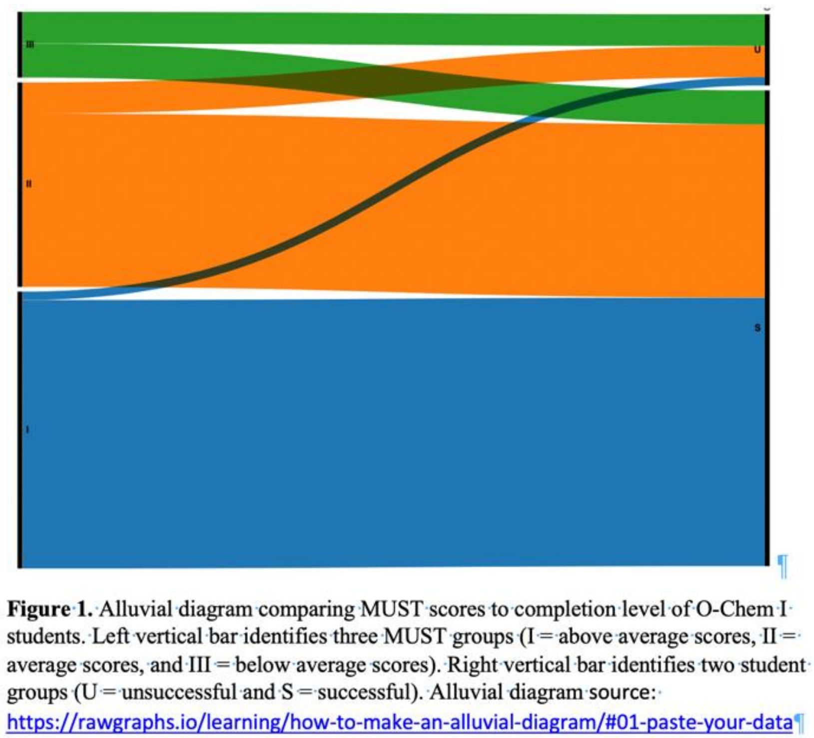


Figure 7. O-Chem MUST ranges and course success [4] (open-source, reprinted with authors' permission).

Why are students in O-Chem, who have completed most of the courses in mathematics needed for their degrees and two semesters of general chemistry still not performing well on the MUST and subsequently not doing well in O-Chem. Regardless of how the data are evaluated what is consistent is that many of the unsuccessful students in Chem I, Chem II, and O-Chem I, began the course with low MUST scores. As stated at the beginning of this paper, the MUST is not just assessing arithmetic skills as was originally thought. Many of the exercises on the MUST require structured established procedures to be solved that go beyond simple arithmetic knowledge. How do these exercises that require procedures to be solved relate to the skills needed to solve O-Chem problems? Both are testing students' ability to learn procedures. Yes, different procedures but still to recall needed procedures. In O-Chem the procedure may be as simple as acid + alcohol → ester + water, but yet this needs to be an overlearned procedure.

Since the NSA team began this project with Chem I and II and now have extended our studies to include O-Chem, our expectations have evolved. The current insights are that the MUST also assesses students' ability to follow known procedures. For example, to solve Q2 on the MUST, the student multiplies (0.50×10^{-6}) (6.4×10^{21}). It is not a given that students know the procedure of how to solve this exercise. Students may be aware that they can take ½ of 6.4 but do they remember that when multiplying base-10 values with exponents, all you need to do is add the exponents? Some may try to take these numbers out of exponential notation, do the multiplication and then put the result back in exponential notation. It is obviously still possible to get the correct answer, but the latter process takes additional time and provides a greater opportunity to make a mistake.

Research Question #2

To what extent does the MUST predict which students will have a satisfactory course average (69.5% or higher)?

Predictability

The NSA team has published three studies that have use the LASSO regression method to determine the predictability of the MUST in determining students' final course success possibility [7,10,14]. LASSO suggests that a 2/3 random sample be used and if possible, then a stratified random sample making sure that all categories evaluated are represented before the predictability values are generated. A practical description of being stratified is when the statistician makes sure that each university is represented by a random sample of 2/3 of each student body. The purpose is to predict students' success (i.e., a course average of ≥ 69.5% or a grade of A, B, and C) or failure (i.e., a course average of below 69.5% or a grade of D and F). All categorial demographic variables (e.g., gender, university level or classification, academic major, ethnicity, first-generation status, and work hours) are combined with the numeric MUST score (0-20 points) to determine the predictability of the

MUST. It was determined in all cases that the MUST has good predictability for both numerical course average (linear regression) and the binary successful vs. unsuccessful (logistic regression).

The statistical modeling using the demographic variables mentioned above provided the following accuracies: 78% in Chem I, 83% in Chem II, and in O-Chem the MUST alone had a 64% accuracy but when the demographic variables of each student's entering science-course GPA (grade point average), and the score on their first exam were added to the equation the predictability rose to 97.0% for the successful student (grades of A, B, or C) and to 82.2% for those predicted to be unsuccessful (grades of D or F) [7]. In all studies, the student's MUST score had a highly significant effect and was a dominant covariate to the overall predictability. The effect of the MUST score is bolstered by the inclusion of other predictive factors, but by itself alone is one of the most influential positive contributors. Another positive contributor to the predictability was when students who are from families where parents and grandparents held degrees, their course success was improved. The variable that appeared to be doing the most harm was when students need to work for over 30 hours/week. However, working on-campus for 1-10 h appeared to have a small positive influence on course success.

Overall, the MUST data are consistent with large effect sizes and repeatable, strong internal consistency. The linear relationship between the MUST scores and course success (Table 2) is just one of many examples that reflect the same trend: the higher the student's entering MUST score the better the chance for student's success in Chem I, II and O-Chem. Keep in mind that these data are combined from multiple universities with varied demographics (public or private institutions; large, medium, and small universities; Hispanic-serving (> 25% Hispanic ethnicity) or Hispanic-emerging (16-24% Hispanic ethnicity) institutions; R1, R2 (Carnegie classification of research 1 or 2 status); and located over a large area of the state of Texas. Even to this day, this CER team still marvels at how a 15-min, 20-Q assessment given to students the first week of classes in varied lower-division chemistry courses can tell us so much about the students in front of us.

RESULTS AND DISCUSSION

Limitations

The results obtained from these data are consistent, reliable, and have provided excellent predictability results requiring no more than 30 min. of class time when taking into consideration the time needed to distribute, explain, sign, and collect the paperwork. Since the MUST is hand-graded, the time needed to score is dependent upon the number of students participating. Students must also pay attention to the instructions and provide the answer that is appropriate, not an alternative correct answer. For example, if asked to give the answer as a decimal number, then that is the only accepted answer (an equivalent fraction is counted as wrong). What needs to be done now is to see if other institutions can obtain similar results.

CONCLUSIONS

The MUST is not limited to assessing the four basic arithmetic operations (addition, subtraction, multiplication, and division). It goes much deeper and identifies students who are proficient in solving exercises that are based on known procedures. The importance of using this valid, reliable diagnostic assessments is that we can identify students who have the potential to struggle with lower-level chemistry courses and identify these students early in the semester when time is still available to institute one or more corrective measures on how to master the procedures needed to succeed. Removing students trusted calculating devices to impress upon them the value of automaticity is a possible start to helping students with low MUST scores. An incentive to remove calculators from the general chemistry classroom is that the MCAT (medical school entrance exam) is a calculator-free entrance exam (over 50% of these students plan to enter the health professions) and this exercise will give them some needed mental-math practice.

In the NSA team's zeal to impress upon the general chemistry community that students' dependence on a calculating device may be hindering their ability to routinely use and make sense of quantitative information [16] in an increasingly data-driven world, we neglected to consider a less intensive mathematics course, like O-Chem. Quantitative reasoning matters in almost every discipline and in every adult role: worker, citizen, and family member [17]. Even the college-educated often lack an understanding of how to make sense of numerical information [18]. Today's people need to be quantitatively literate, (i.e., they need to be able to process and understand quantitative information) [19]. "We need to encourage our students to put aside their calculators and associated cyborgian-thinking patterns, so they can surpass their calculators' capabilities and learn to think conceptually and creatively about quantitative chemistry" [20, p. 730].

We have also learned that students who have committed certain procedure to their long-term memories, may have an easier time with committing other procedures like those needed to succeed in organic chemistry to memory so that these too can be more easily recalled. Success on the MUST goes beyond basic arithmetic understanding and reflects what students have "overlearned." When students possess an ability to retain certain facts in long-term memory this human quality provides an edge to succeed in O-Chem [4]. It is with these types of results from the 15-min. MUST assessment combined with selected demographics that this instrument has proven to be a very valuable tool for identification of at-risk students at a time (first week of class) when it is still possible to provide students with an intervention that may serve to improve their course success.

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. Rocconi, L. M.; Lambert, A. D.; McCormick, A. C.; Sarraf, S. A. (2013) Making college count: an examination of quantitative reasoning activities in higher education, *Numeracy*, 6(2), Article 10.
17. Madison, B. L.; Steen, L. A. (Eds.). (2003). *Quantitative Literacy: Why numeracy matters for schools and colleges*. Woodrow Wilson National Foundation.
18. Cohen, P. C. (2003). Democracy and the numerate citizen: quantitative literacy in historical perspective. *Quantitative literacy: Why numeracy matters for schools and colleges*, 7-20. Madison, B. L., & Steen, L. A. (Eds.).
19. Shavelson, R. J. (2008). *Reflections on quantitative reasoning: An assessment perspective. calculation vs. context: quantitative literacy and its implications for teacher education*, 27-47.
20. Leopold, D. G.; Edgar, B. (2008). Degree of mathematics fluency and success in second-semester introductory chemistry. *Journal of Chemical Education*, *85*(5), 724-731.

ACKNOWLEDGEMENTS

The author thanks all CER instructors for their participation in the Networking for Science Advancement team. In addition, thanks are extended to the ACRICE 5 organization team with special appreciation to Professors Fahmy and Bassioni, and to Ain Shams University for hosting the conference.