**Diagnostic Tests and Student Outcomes:**

**How Important is a Good Chemistry Foundation for General Chemistry?**

Abner Mintz

Assistant Professor, Montgomery College, Rockville Campus

 With many thanks to the important contributions by innumerable professors and staff at Montgomery College who have continued supporting this project semester after semester.

 **Summary: A diagnostic test for first-semester general chemistry students was written and tested by Montgomery College’s chemistry departments. Collection and analysis of the data was originally unsuccessful, but more careful analysis showed that the diagnostic test was actually quite valuable in predicting the odds of student success. The test has since been applied to a second class, a specialized general chemistry class only for engineers, and shown to be similarly valuable but with a different cut-off score for various odds of student success. This article discusses these various results and possible reasons for the differences for different analytical methods and for different student populations.**

**Ancient History (the Rise and Fall of the Diagnostic Empire)**

 Over a dozen years ago, Professor Bob Coley had the inspiration to test the conventional wisdom that a strong chemistry background was important to success in first-semester general chemistry classes (CH 101). Many of us felt that weak students should be encouraged to take CH 099 (a remedial class that brought the students up to speed in high school chemistry before they went on to college-level work) instead, only then proceeding to CH 101. Unfortunately grades in high school chemistry classes proved unreliable in telling a weak student from a strong one, so the advice given to students was highly subjective. Bob Coley wrote a diagnostic test full of questions testing basic chemistry knowledge. The test is multiple choice and designed so that any recent chemistry class should allow a student to do reasonably well (the questions are deliberately not very difficult).

Some example questions:

 What is the name of the element whose symbol is shown below?

 Ni a) nitrogen b) nickel c) sodium d) niobium

What is the density of a sample of wood whose mass is 736 g and whose volume is 957 cm3?

a) 0.769 g/cm3 b) 1.30 g/cm3 c) 221 g/cm3 d) 736 g/cm3

 The test consists of 27 questions. Twelve of the questions are on nomenclature and are worth 1 point each (the first question is an example of a 1-point question); the other 15 are on various other topics such as density, conversions, balancing equations, stoichiometry etc. and are worth two points each, for a total of 42 points possible. The students are given an information sheet with a periodic table on one side. On the other side of the information sheet are various metric prefixes, conversion factors, and formulas (such as the formula for density). This means that a student who has forgotten how many centimeters there were in an inch can still solve a conversion problem involving those units as long as they know what “1 inch = 2.54 cm” means and how to use it. The students have 30 minutes to complete the test on the first day of class.

 Once Bob Coley designed the diagnostic test, it was important to evaluate it. The test was first tried out on a small scale on the students in three CH 101 labs I was teaching that semester. When we looked over the results at the end of the semester, they looked promising: most of the students who had done well in the class had had high scores on the diagnostic and vice versa for those who had done poorly. As a result, the departments of all three campuses agreed to test the diagnostic test on all of their CH 101 classes and report back the data for analysis. It would take several semesters to get enough data for analysis to be highly significant. After several semesters of date were collected and analyzed, it was announced that no correlation at all was found between success on the diagnostic test and success in CH 101. As a result, two of the three campuses (Takoma Park and Germantown) stopped giving the diagnostic entirely. Rockville continued with the test for a simple reason: they found an alternate use for it. The test was divided into sections by topic, so it could still be used to find which topics a given student was weak in and give that feedback to the students. A student who was weak in stoichiometry would discover that at the beginning of the semester and could read ahead and/or pay especially close attention when those sections were covered early in the semester. And if a student who was very weak across the board saw the results and came to discuss them with their teacher, perhaps switching to CH 099 and becoming better prepared for CH 101 as a result, well, no harm done.

**The Middle Ages (the Phoenix Rises from the Ashes)**

 This was the status of the diagnostic test when I finally achieved my long-time dream of gaining a full-time position at Montgomery College. When I was settling in and preparing to teach my first CH 101 class after my return, one of the other professors explained to me how the diagnostic test was being used and why. At the end of my first semester back, I looked over the diagnostic scored in my sections and received a shock – there was a clear match (to my eyes) between the diagnostic score at the beginning of the semester and the results at the end of the semester. I then discussed the issue with several other professors and found that they had seen the same thing; indeed, some of them were using the scores to advise students to switch to CH 099 at the beginning of the semester. Their usual cut-off score where they started getting worried about a student’s performance was somewhere around a 25 out of 42. After mulling this for a while, I went around and begged a number of the other professors to give me their diagnostic data for as many semesters as they had available, pooled it together, and did an informal analysis. The results were clear:

8 sections, 238 students:

Score NA 11-15 16-20 21-25 26-30 31-35 36-40 41-42

%A-C 25% 8% 4% 39% 31% 63% 74% 100%

“NA” stood for the students who managed to avoid taking the test (generally by not being there on the first day of class). For each category I calculated the percent who got grades of A, B, or C (which are the grades I considered to be a successful outcome in CH 101).

 While the very small sample size and possible selection bias from only professors who had saved their data made the results more qualitative than quantitative, the qualitative results were still clear: students with a higher diagnostic score had a much higher chance of getting a decent grade in CH 101. Based on this, I consulted Bob Brenneman (then the department chair) who agreed with me that the issue deserved further investigation. I contacted Ken Weiner, a professor in the math department who was deeply involved in Outcomes Assessment at Montgomery College. He agreed to help collect the data from all professors at the Rockville campus, and we have done so for many years (the current data has up through Fall 2010 results in it). While the numbers from the initial sample above did indeed turn out to be a bit inaccurate (due to the small sample size and collection bias), the official data collected ever since has resulted in qualitatively similar results in every set collected. The overall results can be summarized as follows (the analysis has gotten a lot more detailed over time):

This records the results for CH101 Assessment Tests and the associated grade for students for CH101 for 3084 students over several years.

Range %A-C %A-D %F %W # students

38-42 75% 80% 8% 12% 257

33-37 67% 74% 12% 14% 419

28-32 53% 64% 14% 22% 705

23-27 43% 56% 16% 27% 614

18-22 34% 46% 17% 37% 437

13-17 31% 37% 13% 50% 159

0-12 6% 21% 15% 65% 34

NA 48% 59% 16% 25% 459

 A-C A-D F W

33-42 range: 70% 76% 10% 13% 676

23-32 range: 48% 61% 15% 24% 1319

0-22 range: 32% 43% 16% 42% 630

All (w/o NA): 50% 60% 14% 26% 2625

All (w NA): 50% 60% 14% 26% 3084

 Interestingly this data matches up with what many professors have thought for some time: the overall passing rate (A-D) for CH101 students at our campus is 60% according to this data, which is reasonably close to the 50% retention rate many professors mention as a rule of thumb. Around ¾ of the students with high scores do well in the class (A-C), around ½ the students with midrange scores do well, and only about 1/3 of the students with low scores do well. For the lowest scores, only about 1/20 of the students do well.

 Furthermore, the teachers’ intuition that a score of around 25 was where a student needed advice had some basis to it: it is around a score of 25 or so that the student’s odds of passing drop to around 50/50. A student with an extremely low score is essentially wasting their time and money, and would be much better served taking CH 099 instead and then coming back with the appropriate background knowledge that would allow them to have a reasonable chance in the class.

**Incompatible Results (How Could It Fail and then Succeed?)**

 Once our analysis of the data showed solid evidence that the test was actually useful, there was a major problem to discuss: How could the first analysis, done years ago, show no correlation with success, and our results show a strong correlation? What was going on here? That is a question that may remain a mystery forever. It turns out that after the first analysis showed it to be a failure, the project was not only abandoned but also erased. The data was not kept. The analysis of the data was not kept. While we were able to contact Bob Coley, he didn’t have either data or analysis nor did he remember exactly how the analysis was done. Neither did anyone else – it had dropped out of institutional memory. As a result, we will never be certain how that conclusion was made. I do, however, have a strong suspicion of how it could have happened based on my own analysis: I think they may have only looked at failure rates (% of students who got an F). If you look at the %F, there is almost no correlation between that percentage and the score. The diagnostic test does a very poor job of predicting who will get an F. In large part this is because students who know that they are going to get an F in a class have another option: they may withdraw up until a certain deadline (usually quite late in the semester at Montgomery College). While students may withdraw for any reason even if they are doing well (even a good student can have to drop a class due to some sort of personal disaster or live-changing event), there is little doubt that a student who is doing poorly is much more likely to drop the class to avoid an F on their transcript.



 While the relationship between score on the diagnostic and clustered chances of not passing the class are not precisely linear (they tend to level off at high scores), it is close enough to be interesting (the R2 values for the line for % W+F is 0.9505). I can’t think of any theoretical reason why it should be linear but it is very close.

 If the prior analysis looked only at %F and ignored students who withdrew from the class, it would have been very easy for them to come to the false conclusion that the diagnostic test had no correlation with success in the class.

 Bob Brenneman and a member of the mathematics department did a more detailed analysis of the data several semesters back involving an ANOVA (analysis of variance). While the mechanics of this analysis were too detailed to go into here, some of the results were worth mentioning: 89% of the variation of the percentage of students earning a grade of A-C was explained by their scores on the diagnostic exam. Why was it not 100% predictive? My explanation is that having a good chemistry background was necessary to get a good grade but not sufficient. For example, I have had students who got a score of 42 on the diagnostic test on the first day, indicating that they had an excellent chemistry background. I never saw them again and they never dropped, so they received an F. The diagnostic test explained 79% of the variation in the percentage of students withdrawing from the class. Dr. Brenneman also found that the odds of the data correlation for student success (A-C) being due to random chance was calculated as 7.75 x 10-16%, which is essentially zero; in other words, we can be essentially certain that this is a real effect.

**The Modern Era (What are We Doing Now?)**

 Over time, the strength of the data convinced the various members of the chemistry departments on all three campuses that the diagnostic test was very useful as a predictor. The problem is that students are taking it on the first day of class when it is often too late to rearrange their schedule. We have thus spent several years working on getting the test accepted by the college administration as a placement test. This semester the curriculum committee accepted a change to the prerequisites for CH 101 so that instead of passing CH 099 or high school chemistry, students would have to pass CH 099 or the placement test in order to register for CH 101. The placement test is being turned into a computer program so that it gives similar questions from a test bank instead of always the same questions in the same order on paper. The estimated starting semester is Summer I, 2012. Once the placement test is implemented many of the weakest students currently in CH 101 will be redirected into CH 099, where they will get a stronger chemistry background and be much better prepared for CH 101. We chose a score of 21 for the prerequisite (50% on the test), but have the option to raise or lower that score if 50% turns out to be suboptimal when we analyze future data.

 We are putting into place a new, non-credit class (CH 090) which will consist of one week of intensive training in basic chemistry, 3 hours each day, for students who have already had basic chemistry but are concerned that they may have forgotten much of it. The class should serve as a reminder so that those students can then go on to take the Placement Test and hopefully succeed in being placed into CH 101.

 Another interesting development is that several semesters ago we started giving the diagnostic test to the students in CH 135, a special chemistry class for engineers which involves almost all of the material for CH 101 and CH 102 in one semester (going twice as fast as the material is normally covered). Only the strong science and math backgrounds of the typical engineering student make CH 135 even remotely a reasonable class for the engineering students to take, and even then we lost many of the weaker students each semester. Since CH 135 is a ‘gateway class’ for the University of Maryland at College Park’s engineering program, failing the class or withdrawing from it can have a strong negative impact on an engineering student’s hopes for transferring to UMCP. It would be far better if the weaker engineers were encouraged to take CH 101 and CH 102 separately, or even CH 099 if necessary. The results so far are promising:

This records the results for 634 CH135 students taking the CH101 Diagnostic Test:

Range %A-C %A-D %F %W # students

38-42 89 % 91 % 5 % 4 % 79

33-37 74 % 78 % 8 % 14 % 143

28-32 64 % 72 % 9 % 19 % 129

23-27 53 % 67 % 9 % 24 % 116

18-22 43 % 57 % 13 % 31 % 72

13-17 44 % 50 % 13% 38 % 16 (ss)

0-12 25 % 25 % 13 % 63 % 8 (ss)

NA 46 % 59 % 21 % 20 % 71

 A-C A-D F W

33-42 range: 79 % 83 % 7 % 12 % 222

23-32 range: 59 % 70 % 9 % 22 % 245

0-22 range: 42 % 53 % 13 % 34 % 96

All (no NA): 64 % 72 % 9 % 19 % 563

All (w/ NA): 62 % 71 % 9 % 19 % 634

Percentage of CH 101 (Series 1) and CH 135 (Series 2) students receiving a grade of A-C for each score category on the Placement Test:

 Interestingly, the CH 135 students do better than the CH 101 students despite taking an objectively harder course containing nearly twice as much material; anecdotally the CH 135 professors agree that this is because of the engineering students’ skills in math and science subjects being better, on the average, than those of the CH 101 student population. Though the results are not yet as solid as the CH 101 results due to not having collected as many students’ worth of data, it looks like the test still works for this different population, though a higher cutoff may be possible (or the same cutoff could be used and fewer students shifted to the lower level class).

 In summary, the Placement test for CH 101 has been validated and shown to be very valuable in predicting which students have insufficient chemistry background to succeed in basic chemistry. We are implementing the test in Summer 1, 2012 with high hopes of increasing retention in CH 101 and helping students attend the appropriate class for their current skill level. We are also putting a support system into place for students whose skills are just slightly short of what is needed, and examining the possibilities of extending the Placement Test to a similar but more rigorous class for engineers.