

Beyond Standardization: Accountability in a Project Based Model

Across the country, the controversy surrounding the nature of standardized testing is rearing its head once again. With testing hours being extended, frustrated parents pulling their students out of testing, administrators unsure how to best advocate for their schools, and a student body that is stuck in the middle, it begs the question: What is the proper course of action? To opt out? Or is testing is the best available option? Are there other things we can focus on that are more beneficial than test scores? How influential is pure luck and probability on a student's test scores?

This struggle comes in the wake of the introduction of the new PARCC tests, a multi-state, collaborative effort to craft a tool for measuring the new common core standards and a more accurate measure of student and school progress. While the PARCC test is mostly marketed for grades K-12, the test is creeping its way into the college admissions process. According to the PARCC website, all of the participating states, including Colorado, have made an interesting commitment: *"The Colorado Department of Higher Education, representing both two- and four-year public and private colleges and universities, has committed to participate in PARCC, help develop the college-ready assessments, and, ultimately, use those assessments as one indicator of students' readiness for entry-level, credit-bearing college courses."*

It is certainly reasonable to question if measuring schools and students with standardized tests is the best assessment of their quality. We addressed whether these tests even a reliable representation of individual knowledge. The testing enforcers claim that it does.

To answer this question, we quantified the variation in an individual's ACT score (known as the Standard Error of Measurement - or SEM). We began by looking at specific scores from the graduating class of 2014. Due to the nature of multiple choice tests, students have a chance at getting questions that they do not truly know the answer to right. In order to account for this we ran a Binomial Probability Distribution to find the standard deviation for each score. A 95% confidence interval was calculated with a range of 1 to 3.5 (average of 2); two equivalent students could score as much as 7 points apart (calculations attached [here](#)). From this we know that if two students were to score within a few points of one another it is possible that the higher scorer is better, but it is also reasonable to assume that they are equal. We cannot draw firm conclusions from these scores, only a general sense unless scores are far apart.

We were not able to do a similar analyses for PARCC, both the PARCC and TCAP use a different scoring system that is still far more complex and unknown to us. This scoring system is called IRT and its main purpose is to reduce the guessing factor in each of their tests without increasing the test length. However, we can predict that it is less precise than the ACT based on the lack of a personal incentive by students. We confirmed this by creating a survey

for Animas High School students , to learn how much effort the students put into the standardized tests that they have taken in the past. Out of 128 high school students across the grade levels, the responses that we received reinforced our estimations. The survey results showed that 70% of the students reported trying a fair amount or a lot on the ACT test, with only 30% for the TCAP (the forerunner of the PARCC). Also, 30% of the students reported not trying at all on the TCAP, while only 20% didn't try on the ACT. From our research on the questions on the ACT, it is clear that this test is an imprecise measure of student knowledge. After seeing the results from the student survey, it is safe to assume that the amount of guessing involved with the other standardized tests can skew the results even more than in the ACT, rendering them even less precise.

In efforts to gain better understanding of the precise and complex nature of a standardized assessment, we decided to create our own. Naturally, we needed to find a subject for the test that would encompass something other than educational curriculum. We ultimately decided to focus on something that students at Animas High School would be adequately familiar with- the curriculum and history of Animas High School. We then drafted four different standards that were suitably vague but that students should, theoretically, be familiar with. With the focus of the test determined, we proceeded to write learning targets, or an expectation of knowledge, for each. The final content categories were: Mission & Vision, Culture, Structures, and History.

Ten questions were written to address the standards of each content category. All of the questions were intentionally written with the expectation of roughly 40-60% of students being able to answer them correctly. To determine questions that are answered correctly by our target percentages of students, we gathered baseline data from three classes. Based on this data, questions that too many students answered correctly were cut from the test. Questions that were answered incorrectly by the majority of students were cut as well; even if the question itself was labeled relevant to our schools' curriculum by our initial draftings. Questions were then selected from their content category by their quality, and placed into a final pool of ten multiple choice questions, finalizing our Animas Standardized Assessment.

After our student body took the test, we devised a scoring system. Students were split into four categories: advanced, proficient, partially proficient, and failing. To decide what scores fit each category, we ran a normal distribution on the scores, and set ranges to encapsulate the following percentages - 10% advanced, 50% proficient, 30% partially proficient, 10% failing. The students scores were established based not on individual performance, but on their ranking within the data set. Doing this enabled us to categorized all answers into in certain, pre-determined percentages.

As a class, we hope that our cultivated knowledge can help us inform the school position on assessment. We accept that it is an ongoing and important subject, and would hope the school plans for future standardized assessments with as much knowledge as possible.

