

The Board set the standards for the SOL tests on the first form constructed of each test. These cuts were selected on the number correct raw score. Now, if this were the only form of the test ever used, the cuts would always be the same. But, since we can't continue to give the same form over and over again, other forms of the test must be developed. Herein, lies the problem.

Each new form of the test is developed with the intention that it is the same difficulty level as the first form constructed (the forms that the standard was set on). Several statistical methods are employed during form construction to try to make the forms equal in difficulty to the first form constructed. But, let us say for the sake of this discussion, that the new form that was constructed was slightly more difficult than the first form constructed. If this were the case, would it be fair to expect a student to achieve the same raw score as the students that took the easier first form constructed? The answer of course is, no. Conversely, if the new form was slightly easier, it would not be fair to the previous students that the students taking the easier form should only need to get as many right as they did on the harder test. Scaling and equating are the tools we use to make sure that each student receives a fair and equitable score on the test.

The method (equating and scaling) we employ to create the equated scale scores are used by all high stakes testing programs. For example, the raw score necessary to achieve a 600 on the SAT mathematics test varies across administrations due to equating and scaling.