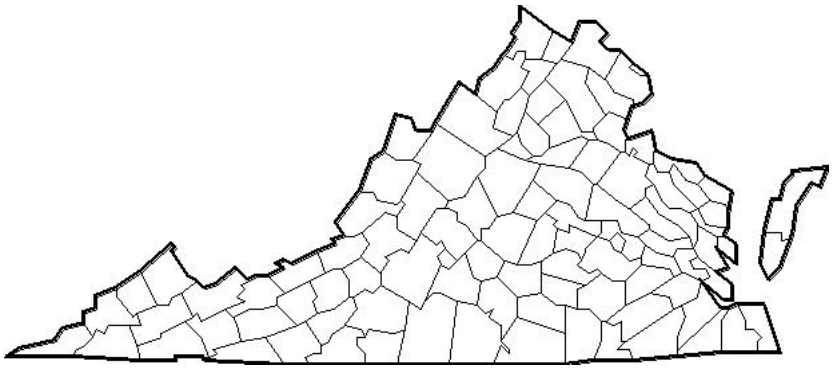
A paper presented at the April 2004 meeting of the American Educational Research Association San Diego, CA

A Validity and Calibration Study for a Set of Standards Diagnostic Tests



TESTS FOR HIGHER STANDARDS

David E. W. Mott and S. Stuart Flanagan

A Validity and Calibration Study for a Set of Standards Diagnostic Tests

David E. W. Mott and S. Stuart Flanagan

Abstract

Test for Higher Standards publishes diagnostic tests for sale to districts and schools in states with high-stakes assessment programs. This study had two very practical purposes. The first purpose was to correlate / validate 22 of our *Simulation Tests* (SIMs) to the equivalent *Virginia Standards of Learning Tests* (VA SOL). The second purpose was to calibrate each of our (SIMs) to the equivalent VA SOL test. We were not trying to do a precise equating between the two; rather, we were trying to give our clients general guidance about the difficulty of our tests compared with the state tests. We also sought to establish some critical score ranges for instructional guidance.

A volunteer sample of classrooms in client schools and districts was selected. For each student, the teachers recorded pairs of test scores: SIM test number-of-items correct and VA SOL test scaled-scores (later converted to number-of-items-correct). Both set were converted to percent correct.

The first correlate/validate step was accomplished, the correlation coefficients ranged from .60 to .85, average = .72, for *n*'s from 91 to 557, average *n* = 294, total *n* = 6,468. The second purpose was to calibrate each of our *SIM Tests* to the equivalent *VA SOL Test*. We found that, out of the 22 tests, all of our tests were harder than the equivalent state tests except: Mathematics, Grade 3; History and Social Science, Grade 8; and World History I. The state tests were slightly more difficult for these three.

Finally, cut points were established for each of our tests to divide the scores into three score Action Ranges. A lower range of scores where most students who scored in this range a month before the state tests were unlikely to pass. Instruction for these students probably needs to focus on teaching precursor skills. The next was a middle range where the outcome of the state test was much in doubt. The majority of the standard-specific remediation likely needs to be focused on students in this range. Above the second cut point, we expect most of the students to pass the state test without much difficulty. Students in the top Action Range should be remediated on identified weaknesses, but, beyond that, can be given primarily enrichment activities.

A Validity and Calibration Study for a Set of Standards Diagnostic Tests

David E. W. Mott and S. Stuart Flanagan
Tests for Higher Standards

When *Tests for Higher Standards* first published its Virginia Standards of Learning (SOL) diagnostic tests in 1998, we were primarily concerned with their content validity. We wanted to be sure that the test questions measured specific Standards (referred to hereafter as SOLs). We were also intent on making our tests, taken as a whole, reflect the entire tested portion of the standards in each grade and subject. A blueprint, modeled after the appropriate *Virginia SOL Test* blueprint, was created for each *TfHS* test so that a balanced set of items could be written. Because the primary use of the tests was to be diagnostic, we needed enough questions on each test to be able to indicate individual student's strengths and weaknesses, right down to a specific SOL. The full content of the Virginia Standards of learning is available on the VADOE Website (VA DOE, 2003).

Our first step in ensuring validity was to be sure that each question was written with deliberate attention to the meaning and specific wording of the SOL to be measured, taken in the context of the rest of the SOLs. Each question was written to measure one specific SOL. Then, test authors crosschecked each other's questions to ensure each one was correctly classified and well written. Later, classroom teachers and subject-matter experts reviewed questions for SOL match, for clarity, and for appropriate content. Changes and corrections were made as appropriate. As soon as tests were developed, they were administered to several classrooms of students as a pilot to see if students correctly understood the questions. We also tried to make sure the questions were of appropriate difficulty. Tests were corrected again. When this study was conducted, the tests have been through several rounds of pilot and field-testing. We believe all of these reviews, including the continual ongoing review by our users — teachers and students, resulted in tests which have a high degree of content or face validity.

However, we were also interested in measuring what is usually called the *predictive validity* of our tests against the State's SOL Test. In the present study, we are reporting on what might be described as the predictive validity of each of the *TfHS Simulation Tests* in reference to the state's equivalent test. We have several sets of tests, designed for different uses.

We chose to use our *Simulation Tests* in this study because they are designed to be as close a match as possible to the *State SOL Tests.* Scores on these tests would be expected to correlate most highly with scores on the state test. For these tests, the state's test blueprint is followed in detail. We use the same typeface as the state test and also use the alternating "A B C D" and "F G H J" labels for the question distracters. We should also note here that the simulation tests **are not diagnostic to the SOL level**, but only to the more general level of "Reporting Categories" (roughly, strands) that *Virginia SOL Tests* use. The simulation tests are ten items longer than the state's blueprint shows, because the state administers ten experimental questions. We wanted the

students' experience in taking the *Simulation Tests* to be similar to their state test experience, which dictated that the length of our tests needed to be equal to that of the state tests. Therefore, we distributed the ten extra questions across the eligible content. There are *TfHS Simulation Test – Virginia SOL Test* pairs in the five subject areas: Reading, Writing Mathematics, Science, and History and Social Science, at grades 3, 5, and 8 and for secondary courses — 25 pairs in all. Only 22 pairs are reported here. (See Results and Discussion, Column 1 below.)

We sought to obtain predictive validity coefficients, which are the product-moment correlations of a student's score on one test matched with that same student's score on a second test. We also wanted to calibrate the difficulty of each *TfHS Simulation Test* to the equivalent *Virginia SOL Test.* We wanted to be able to recommend appropriate "action points" for our tests. We choose to concentrate on "action points", rather than passing levels, because our tests are **not** designed to function as gatekeepers, but rather as a diagnostic tool for teachers and administrators.

Later in this report, we also refer to the *TfHS Grade Level Tests,* which were the first tests we developed. They generally have from three to five questions measuring *each* SOL (or sub-SOL bullet, "•"). These survey tests enable teachers to examine students' strengths and weaknesses down to a specific teachable unit and are designed to be given as intact tests (in one or two sittings). They are longer and give more complete coverage of the Standards than do the *Simulation Tests* but are less like the *Virginia SOL Tests*.

We would like to emphasize here that this is applied research. As a company, we needed answers. Our clients, teachers and administrators in districts using our products, needed answers. We have tried to give the best answers possible, given our data. Sometimes we need to produce answers on data that made us as researchers uncomfortable in its sparseness. People will set "passing scores" on tests, given no student response data at all. (Generally, 70 or 75% is called "passing".) Therefore where we could make even an educated guess, we have attempted to do so.

Method

The study we designed was quite simple. Participating teachers were those in schools scheduled to give the *TfHS Simulation Tests* toward the end of the school year, about one month before the *Virginia SOL Tests* were to be administered. The sample was in no way random, but was a sample of convenience. We only asked teachers in districts where our contact person was willing to sponsor the study. All districts regularly used our tests. Participants were to record pairs of student scores on a data collection sheet. They were to record the number-correct scores for the simulation test and the scaled score for the State test. They were to not to record the students' names, but they were to insure that each pair of scores represented a single student.

Scaled scores were converted into number-correct scores for use in most of the calculations done using tables supplied by the Virginia Department of Education (VDOE). (Our thanks to Dr. Robert Triscari at the VDOE for supplying these tables.) We used number correct scores in calculating correlation statistics since both scores in each pair were more likely to have similar distributions. (We did test this, but the results are not presented here.)

The pairs of test were to be calibrated at the State passing cut-point using an equal percentile method. The number-correct score on the *Virginia SOL Test* equivalent to the passing score was looked up in the VDOE tables. The percentage of students *in our sample* scoring at this score or lower was calculated for the *Virginia SOL Test*. From this, the equivalent cumulative-percent score on the *TfHS Simulation Test* was found. This score was designated the "passing score" on the Simulation test. We also calculated two critical points on the test score scale. The lower critical point was set at a scaled score of 350 and the upper point at 450. These are symmetrical points, equally distant from the passing score of 400. The State scaled scores form an approximately equal interval scale. We picked these points to mark out a range on the *Virginia SOL Test's* scaled scores within which there was doubt in the outcome. The standard error of measurement (SEM) at the passing score for the 1999 Virginia SOL Tests, averaged over 22 tests, was 17, with a range from 13 to 24 (VA DOE, 1999). Thus, this 100-point score range is about ± 3 SEMs for the average test. Within this range will be nearly every student for whom there is any doubt about whether they will pass or fail the *Virginia SOL Test*.

Within this score range the presence or absence of good, targeted instruction can make a difference in the outcome — whether the student will indeed pass or fail. We have designated the equivalent calibrated points on the *TfHS Simulation Tests* "Action Points." *It is important to emphasize that these are all probability statements. All tests measure with error; any student's test scores on a given occasion might not be a reflection of that student's true attainment. We can make no prediction about whether a* particular *student will pass or fail. We can only deal with probabilities.* (We should point out that it is relatively more likely for a student to obtain a score that is unrealistically low on our tests than one that is too high. Some students simply will not take the "preliminary" or "simulated" SOL test seriously.)

The two action points divide the number-correct scale into three regions: In the center region, between the two action points, about half of the students are probably going to pass the *Virginia SOL Test*, without intervention. At the same time about half will fail, without any further teacher intervention. Above the upper action point, *most* students will pass the test, even without much more instruction. Below the lower action point, *most* students will probably not pass, even with some reteaching.

Results and Discussion

A total of eight school divisions (districts) participated in the study. Approximately 14 schools were represented. We do not know exactly how many actual classes/sections are represented, because several different classes/sections were often recorded on the same form. In many cases, school or district administrators filled in the forms rather than teachers. However, we estimate that over 250 classes' data are included in the study.

We will use the Table 1 as a guide to understanding what we found. The table has 29 columns and is divided into sub tables a through d for this report. Results were calculated for each of the tests separately and are presented in Table 1 (a – d), *Simulation Tests — Validity and Calibration Study — Results*.

Column 1 of Table 1 contains the name of the *TfHS Simulation Test* that was compared with each *Virginia SOL Test* of the same name. Note that the grades K through 3 Reading tests cover both reading and writing standards. Although the *TfHS Simulation Tests* and the State *SOL Writing Tests* above grade 3, contain both editing type multiple-choice questions and actual writing samples, we were unable to examine any of these writing tests in this study, as the tests are given earlier in the year than this study was conducted. The results for the 22 test pairs studied are presented in the remainder of the table.

Columns 2 and 3 list the number of items each test contained. Note that the *TfHS Simulation Tests* are approximately 10 items longer that the *Virginia SOL Tests*, because each state test contains the 10 experimental items not used to calculate a student's score. To the students taking the tests, the tests are generally the same length, except for grade 3 English, where the simulation test is shorter.

Columns 4 and 5 show how many students in the sample took each of the paired tests. The number of students taking a test ranged from a low of 91 students who took the Earth Science tests to a high of 557 students who took the World History 1 tests. The cumulative results indicate that a total of 6,457 students in our sample took the *Virginia SOL Tests* and 6,589 took the *TfHS Simulation Tests*. An undetermined number of the students' data is contained under more than one test pair, as results are presented separately for each set of tests they took. A few students took just one of a pair of tests. Their partial results were used, were appropriate, to describe the study's student sample.

Columns 6 and 7 indicate the time period between the administrations of the pairs of tests. There is a great deal of missing data here. Although teachers who recorded the data were asked to indicate the dates of administration for both tests, most did not do so. Some only gave the month and year. As we had indicated that the *TfHS Simulation Tests* were to be given between six to two weeks before the *Virginia SOL Tests*, we have assumed that most teachers did comply with this request. Although the average recorded data indicates that the *TfHS Geometry Test* was given six days *after* the State test, we believe this is a data collection anomaly. We believe that the majority of the tests were probably given in the prescribed period, and before the State tests.

Columns 8 through 12 report the difficulties of the tests in terms of the item (question) difficulty. Columns 8 and 9 report this information in terms of the *number* of items answered correctly and columns 10 and 11 report the same data in terms of the *percentage* of items correct. These last usually called the "*p*-values" of the items. For the first test, English R & W, Grade K-3, we see that students in our sample answered an average of 31.9 items correctly, or 70.8% of the 45 items of the State SOL Grade 3 English test. These same students answered 20.5 items correctly, or 58.6% of the 35 items on the equivalent simulation test. From the point of view of the students, the *TfHS* test would seem harder than the State test, as they knew the correct answer to a smaller percentage of the items. Column 12 graphically displays the relative difficulty of each of the pairs of tests from the point of view of the mean percent of the items answered correctly. *Of the 22 test pairs in this study, all but three of the TfHS tests were harder than the Virginia SOL Tests*. The pluses in column 12 indicate the *TfHS* test was harder, for example,

English K-3. Minuses show that the *TfHS* test was easier, for example, Mathematics, Grade K-3.

+ or –	=	< 5% difference
++ or – –	=	5–10% difference
+++	=	10–15% difference
++++	=	>15% difference

Columns 13 through 16 report the calibrated difficulties of the pairs of tests. Column 13 is simply the number-correct score on each *State SOL Test* which corresponds to the passing scaled score of 400. Column 14 is the equal-percentile calibrated equivalent for the *TfHS Simulation Test*. It is the score with the closest cumulative percentile to the score in column 13. *Again, this score is our best estimate of the score on the TfHS Simulation Test closest to the lowest passing score on the Virginia SOL Test.* This could be called the calibrated passing score. Columns 16 and 17 present the data in columns 13 and 14 converted to percentages. This makes it easier to make it easier to compare pairs of tests where the tests have different numbers of items. [The equi-percentile calibration procedure used here is unrelated to fact that both tests are criterion- not norm-referenced. Equi-percentile calibration is simply the curvilinear equating procedure which is best used with tests of these types. We are using the term "calibration" rather than "equating" because we do not wish to imply that the tests are truly equated in a strong psychometric sense.]

Because both the State tests and our tests measure students' achievement with error, we can not say that a given student who attains the passing number correct on the *TfHS* test *will* pass the *Virginia SOL Test* about one month later. What we can say is the following: **The *majority* of students who score below the passing score on the *TfHS Simulation Test* will fail the *State SOL Test* if they receive no further instruction; whereas, the *majority* of the students receiving a score at or above this passing score will pass the** *Virginia* ***SOL Test.*** This conclusion only applies if the conditions under which the students take the tests are reasonably similar to those in this study. For example, if the Simulation tests were given earlier in the school year, we expect more students at the cut point (passing point) to pass the *Virginia SOL Test* because they have had more time to learn the skills. See the limitations section below.

Column 15 reports the percentage of students statewide who passed the *Virginia SOL Test* in the spring of 2000. This percentage is included so that we can compare our samples of students with the statewide population. Comparing columns 15 and 16, it appears that our samples of students scored below the average for the state test for all of our elementary and middle school tests, with the exception of Science, grade 8, where our sample and the state population had essentially the same achievement. On the other hand, our sample scored above the state average for all the end-of-course tests, with the exception of Chemistry, where our sample and the state population were basically equal in achievement. Clearly, the students in classes we sampled in the elementary grades were from a different population than the population sampled for the end-of-course tests. However, the equal-percentile method chosen for calibrating the sets of tests is not too sensitive to such a population effect with paired samples.

Columns 17 through 21 Are concerned with the results calibration procedures. Columns 17 and 18 represent the numbers of students passing the tests. Column 17 shows the actual numbers of students in our sample that passed the state test. The number in column 18 is the number of the same students who passed the *TfHS Simulation Tests* using the passing points listed in column 14. The numbers in columns 17 and 18 are close, but not identical. This is because the equal percentile equating procedures cannot equate more precisely than one percentile point, and less are precise than that if not all possible scores are represented in both samples, as was the case here for many tests. This again represents the probabilistic nature of the data. Skipping column 19 for the moment, columns 20 and 21 present the data in columns 17 and 18 in percentages. This makes it easier to make it easier to the compare passing rates on the two tests where each test-pair had different numbers of students taking it. Again we see that the percentages are very similar, taking into account the "graininess" of the data. Column 19 is presented next to column 20 so that we can compare the achievement level of our sample to that of the entire state. In all of the elementary samples except for the Grade 6-8 Science test, students in our sample were substantially below the state average in their performance. In the Grade 6-8 Science test the students in our sample's performance essentially equivalent to that of the average student in the state. The situation was reversed for the end-of-course. Students in our sample performed well above the state average. In the Biology test, all of the students in our sample passed the State test.

Columns 22 and 23 indicate the validity coefficients (product-moment correlations) between the State SOL Tests and the TfHS Simulation Tests. Column 22 lists the coefficients themselves and column 23 indicates the number of students' paired tests upon which each coefficient is based. The coefficients ranged from a low of 0.619 (Algebra II) to a high of 0.850 (Chemistry). The mean of the 22 coefficients is 0.719. Overall, we believe these coefficients are appropriate for the natures of the two tests and for the times of the year they were given. The correlations would probably have been higher, had the tests been given closer together in time and had the teachers *not* had the feedback about their students' performance that was given by the *TfHS Simulation Tests*. It may also be that some students did not perform to the level of their ability on the *TfHS* tests, as they were not high stakes for the students. This may have lowered the correlations. We hope, and believe, that teachers did use the information provided to instruct students in areas of weakness. We believe that the magnitude of the coefficients shows the following. When two separate groups of competent, professional test authors (*i. e.*, the *TfHS* test authors and the authors of the state tests at *Harcourt Educational Measurement*) create multiple-choice tests to measure given sets of standards, the tests they produce are fairly equivalent. This is very encouraging. In fact, the findings of this study not only validate the *TfHS Simulation Tests*, they also, reflexively, serve to validate the *Virginia SOL Tests*.

Columns 24 through 29 are included because we believe that many of our users have used a score of 70% of the items correct to indicate passing on the *TfHS Simulation Tests* and, by implication, the prediction of passing on the *Virginia SOL Test*. Comparing the numbers in column 21, number of items correct corresponding to 70% correct for the *TfHS Simulation Test* with column 14, the recommended passing number-correct scores, we conclude that 70% is too high a passing score for all these tests, except for Mathematics, grade 3. For that test, 70%

correct is within 1 number-correct point of what we calculated for a calibrated passing score. On the average 70% passing is 16.5% too hard. If we wanted to set one passing score for all of our tests it would be 53% of the items correct. However, we do *not* recommend this. We believe that the calibrated individual passing scores should be used instead.

Scatter Plots. Figures 1 through 6 are six representative scatter plots of the relationships between the number-correct scores on the *TfHS Simulation Test* and the number-correct scores on the corresponding *Virginia State Test*. Figure 1 shows US History. This plot is of 438 students who took both tests. The correlation of 0.647 is slightly below the average for the tests reported here. There is a slight suggestion of greater ceiling effect with the with the *Virginia SOL Test* than with the *TfHS Simulation Test* (note concavity downward). Figure 2 shows the Reading Grade 3 tests. The correlation of 0.748 was based on 302 students' results. Although the simulation test has only 35 items to the 45 on the state test, no real ceiling effect is apparent for either test. Figure 3 shows the one the two highest correlations, 0.849, based on 302 students' results. No ceiling effects here. Figure 4 is the plot of the Earth Science scores. This is our lowest correlation of 0.595 based on only 91 scores. Figure 5 is of the Mathematics Grade 8 scores. Based on 186 students' scores, the correlation was 0.736 and there was some indication of downward concavity, indicating that there might be a ceiling effect on the state test side. Figure 6 shows 156 scores on the high school end-of-course Reading, Literature, and Research tests. The correlation was 0.678 and no sign of ceiling effects for either test.

Limitations of the study. We need to point several limitations to the study and its conclusions.

Time of Testing: If students take the Simulation test more than one month before they take the State test, the look-ups in the tables presented here will probably under-predict the score on the State test. That is, if the students were to take the *TfHS Simulation Test* in December, we would expect their scores to be lower than if they took the same test in March, because they would have had less instruction. This time-to-instruct effect is likely to be especially strong when there is a much completely new content to be learned, such as in Algebra II or Chemistry.

Unmotivated Students: Students need to be motivated to do well on the *TfHS Simulation Test*. Although there are no external stakes on this test for the students, getting a correct diagnosis of strengths and weakness is to the students' benefit. Students need to know this and need to believe the teacher will actually use the data to improve learning. If the students are unmotivated to do well on the Simulation test, their scores will under-predict their performance on the State test. Students presumably will be highly motivated to do well on the *Virginia SOL Test.*

A Small Sample Size: Only 91 students for one school took the Earth Science test, so use the results for that test with caution. For Biology, all of the students in our sample passed the *Virginia SOL Test,* so the calibration process was conducted at a higher % correct score,

Sensitivity to Instruction: A most important point to be made here is simply this. *It is a teacher's job to destroy the correlations between the* Simulation Test *scores and scores on the* Virginia SOL Tests*!* If the teachers use the diagnostic information provided by this test and from other sources well, they can raise the level of learning in their students. Correlations between the sets

of tests will be lower to the extent the teachers have been successful diagnosticians and instructors and the students have been successful learners.

The 1995 History and Social Science SOL Content: The study covers the administration of the *TfHS Simulation Tests* for the 1995 History and Social Science SOLs, coupled with the *Virginia SOL Tests* based on those same standards. Our newer tests based on the 2000 edition of the SOL are not covered, nor are our transition materials based on the state-specified "crosswalks" between the two sets of standards.

Recommendations

We have prepared a table that may be used to determine what strategies teachers should use to instruct students prior to the SOL, based on their overall number-correct score on each *TfHS Simulation Test*. Three Action Ranges are listed for each test. These ranges are set by the data collected in this study. This table is based on the original passing scores established by the State for the History and Social Science Tests (HSS). Table 2a presents similar Action Ranges for HSS tests using *the new passing scores* established by the Virginia Board of Education on November 28, 2001. All other passing scores were unchanged.

Table 2, TfHS Simulation Test Score Action Points for Teachers and Administrators. Use the Action Range scores in deciding how to teach your students. The regions divide the number-correct scale into three regions.

– *In the Center Action Region,* teachers need to concentrate heavily on teaching those skills identified as weaknesses on the test. As many as half of these students will may fail the *Virginia SOL Test* without extra help. Teachers need to provide this extra help. Direct instruction on those areas shown to be weak will be helpful. If necessary, use one of our other tests, another assessment, or your own knowledge of what the students know, to make specific diagnoses. Most of these students probably *can* pass with skill-directed help and general review.

– *In the Upper Action Region*, teachers should be concentrating on broadening, generalizing, and enriching the students' knowledge and understanding of the subject area. Few of these students are likely to fail the *Virginia SOL Test*. Nevertheless, any skill weaknesses that are shown by the *TfHS* tests should be addressed.

– *In the Lower Action Region*, teachers should understand that major remediation and reteaching will be necessary to get the students to pass the *Virginia SOL Test*. Many of these students will not pass, even with this reteaching. However, it is very important that these students be encouraged to take the state test seriously and to do their best. *Some* of these students will pass the *Virginia SOL Test*; more will, if they receive extra help. It may prove useful use one of our other tests, possibly one for an earlier grade (e. g., the fourth-grade *Grade Level Test*, if the student is in the fifth grade) to find specifically where gaps in a student's achievement may be.

An extension to the study. *TfHS* publishes two other series of tests in addition to the *Simulation Tests*. We publish *Grade Level Tests* and a product we call *Student Achievement Booklets*.

TfHS's primary product is our set of diagnostic *Grade Level Tests*. We believe these tests are our most instructionally useful test, as they are available in every grade between kindergarten and grade 11. We publish five tests in four subject areas per grade. (Reading and Writing are separate tests.) These tests are significantly longer than the *TfHS Simulation Tests*. Each SOL is measured and diagnosis is more precise. Diagnosis is at the level of SOL rather than the less-precise diagnosis at the level of reporting category produced by the Simulation tests.

We also publish Students Achievement Booklets, which consist of individual SOL level tests for use as end-of-unit tests or the like. This set is designed to be primarily a set teacher's resource tools for use before, after, or before and after a unit of instruction. We report no data for these tests here.

Table 3 (a-c), TfHS Grade Level Test Score Action Regions for Teachers and Administrators, is an extension of the results from this study to our *Grade Level* series of tests. It is based on an extrapolation of the *Simulation Test* data to the *Grade Level Tests*. As such, it is based on *no* primary *Grade Level Test* data. Because we have *not* calibrated the *Grade Level Tests* to the *Virginia SOL Tests* in any direct way, we are much less secure in our placement of the *Action Regions* in this table. Normally, we would not publish such a table. We have never recommended a "passing score" for any of our tests. We are producing this table at this time only because the *Simulation Test* data allows us to make some *educated guesses*.

The content of the Simulation tests substantially overlaps the content of the *Grade Level Tests*. From carefully looking at the test questions themselves, and from our knowledge of the simulation tests' *Action Regions*, we are producing these tentative *Grade Level Test Action Regions*. We believe that such an educated guess on our part is better than simply saying, "70% (or 75%, or some other arbitrary figure) is passing." We suggest that our clients use our recommendations until they collect their own local data.

References

VA DOE. (1999). Virginia SOL Assessment Technical Report: 1998-1999 Administration Cycle (PDF format). http://www.pen.k12.va.us/VDOE/Assessment/TechReport_98-99.pdf

VA DOE. (2003). Standards of Learning Currently in Effect for Virginia Public Schools. Richmond, VA. http://www.pen.k12.va.us/VDOE/Superintendent/Sols/home.shtml

About the authors

David E. W. Mott and S. Stuart Flanagan are the co-proprietors of Tests for Higher Standards, an educational test publisher based in Richmond, VA and North, VA. David Mott is a psychologist formerly employed at the Virginia Department of Education. Stuart Flanagan is Professor Emeritus, College of William and Mary, School of Education. Correspondence should be addressed to: David E. W. Mott, 2821 Ellwood Ave., Richmond, VA 23221, USA; e-mail demott@tfhs.net.

Table 1a. Simulation Tests — Validity and Calibration Study — Results

	Column 1	*Col. 2*	*Col. 3*	*Col. 4*	*Col. 5*	*Col. 6*	*Col. 7*
		Number of Scored Items Contained in the Test		Number of Students in the This Study Sample		Days Between Test Administrations (estimate) ***	
	Subject Area/ Grade	Virginia SOL Test	TfHS Simulation Test	Virginia SOL Test	TfHS Simulation Test	Days Between	Number of Students for Whom Data is Available
1	English, R & W, Grade K-3	45	35	181	311	39	288
2	English, Read. Lit., & Res., Grade 4-5	42	52	159	160	22	44
3	English, Read. Lit., & Res., Grade 6-8	45	52	185	185	-	-
4	Mathematics, Grade K-3	50	60	438	444	30	99
5	Mathematics, Grade 4-5	50	60	158	162	27	30
6	Mathematics, Grade 6-8	60	70	188	187	-	-
7	Science, Grade, K-3	40	50	468	466	24	157
8	Science, Grade, 4-5	40	50	286	290	21	32
9	Science, Grade, 6-8	50	60	161	161	-	-
10	History and Social Science, Grade K-3	40	50	418	409	22	221
11	History and Social Science, Grade 4-5 **	40	50	287	290	18	24
12	History and Social Science, Grade 6-8 **	50	60	160	160	25	84
13	English, Read. Lit., & Res., High School	42	52	173	168	22	59
14	Algebra I, High School	50	60	410	414	9	131
15	Algebra II, High School	50	60	530	525	9	156
16	Geometry, High School	45	55	504	490	-6	158
17	Earth Science, High School *	50	60	91	91	-	-
18	Biology, High School	50	60	118	116	8	120
19	Chemistry, High School	50	60	216	216	-	-
20	World History I, High School	61	71	557	558	6	142
21	World History II, High School **	63	73	326	327	7	97
22	U. S. History, High School **	61	71	443	459	5	78
	Sums or Averages			*6,457*	*6,589*	*16.9*	*1,920*

* Use with caution, as only of one HS (one teacher, 3 classes) is represented.

** The passing scores were changed by the Boad of Education on 11/27/01. Numbers in () reflect the *NEW* passing scores.

*** Most of the respondants did not supply this information.

Table 1b. Simulation Tests — Validity and Calibration Study — Results

	Col. 8	Col. 9	Col. 10	Col. 11	Col. 12	Col. 13	Col. 14	Col. 15	Col. 16
	Mean Number of Items Correct on the Test (in this sample)		Mean Percent Items Correct (in this sample)		Comparison of the Item Difficulty on the TfHS Simulation Test to the State SOL Test	Number of Items Correct Corresponding to the State Passing Score		Percent of Items Correct Corresponding to the State Passing Score	
	Virginia SOL Test	TfHS Simulation Test	Virginia SOL Test %	TfHS Simulation Test %	(+ Sim Test Harder – Sim Test	Virginia SOL Test	TfHS Simulation Test	Virginia SOL Test %	TfHS Simulation Test %
1	31.9	20.5	70.8	58.6	+++	33	21	73.3	60.0
2	28.8	23.9	68.5	45.9	++++	29	22	69.0	42.3
3	27.4	25.5	65.1	49.1	++++	28	25	66.7	48.1
4	36.3	44.1	60.5	73.5	-	35	43	70.0	71.7
5	31.0	33.8	62.0	56.3	++	33	35	66.0	58.3
6	37.6	31.1	62.7	45.6	++++	34	29	56.7	41.4
7	27.7	27.7	69.2	55.4	++	28	30	70.0	60.0
8	25.7	25.7	64.2	51.4	+++	25	25	62.5	50.0
9	33.6	28.9	67.3	48.1	++++	28	24	56.0	40.0
10	26.4	28.5	66.0	57.1	+	26	31	65.0	62.0
11	25.6	30.0	63.9	60.0	+ (+)	27 (26)	33 (32)	67.5 (65.0)	66.0 (64.0)
12	29.2	32.5	58.3	54.2	- (+)	33 (28)	40 (34)	66.0 (58.0)	66.7 (56.7)
13	31.5	34.9	75.0	67.1	+	22	26	52.4	50.0
14	34.0	36.3	68.0	60.5	+++	28	27	56.0	45.0
15	37.4	37.2	74.8	61.9	+++	28	26	56.0	43.3
16	34.5	35.0	76.6	63.7	++++	26	23	57.8	41.8
17	35.4	29.3	70.8	48.8	++++	29	22	58	36.7
18	41.9	43.5	83.7	72.5	++++	25	21	50.0	35.0
19	30.8	36.2	61.7	60.4	--	24	32	48.0	53.3
20	44.9	51.9	73.7	73.1	+	30	33	49.2	46.5
21	43.8	57.2	69.6	78.4	---	33 (29)	47 (44)	52.4 (46.0)	64.4 (60.3)
22	43.5	42.3	71.4	59.5	+++	42 (36)	38 (34)	68.9 (59.0)	53.5 (47.9)

Table 1c. Simulation Tests — Validity and Calibration Study — Results

	Col. 17	***Col. 18***	***Col. 19***	***Col. 20***	***Col. 21***	***Col. 22***	***Col. 23***
	Number of Students Passing in This Sample		Percent of Students Passing the Virginia SOL Test in 2001, Statewide %	Percent of Students Passing in This Sample		Validity Coefficients (Correlations) Between Scores on State SOL Tests and the Scores on the TfHS Simulation Tests and the Numbers on Which Correlations Are Based	
	Virginia SOL Test	TfHS Simulation Test		Virginia SOL Test %	TfHS Simulation Test %	*R*	*n*
1	181	178	65	57.8	57.2	0.748	302
2	106	92	73	66.7	57.5	0.653	156
3	101	96	73	54.6	51.9	0.666	185
4	277	284	77	63.2	64.0	0.761	429
5	76	78	67	48.1	48.2	0.706	156
6	113	113	68	60.1	60.4	0.763	186
7	263	252	84	56.2	54.1	0.717	460
8	167	164	75	58.4	56.7	0.667	285
9	118	113	74	73.8	70.2	0.798	161
10	175	174	72	41.9	39.4	0.677	395
11	135 (152)	135 (146)	63 (70)	46.9 (52.8)	46.4 (51.2)	0.713	286
12	56 (81)	46 (73)	56 (77)	35.0 (50.6)	32.9 (49.1)	0.683	159
13	161	157	82	93.1	93.5	0.678	156
14	344	349	74	83.9	84.3	0.642	403
15	435	436	74	82.1	83.0	0.619	525
16	448	435	73	88.9	88.8	0.770	486
17	73	75	73	80.2	82.4	0.595	91
18	118	115	81	100.0	100.0	0.636	114
19	159	156	74	73.6	72.2	0.849	216
20	507	503	83	94.1	93.1	0.850	557
21	272 (293)	277 (292)	65 (71)	83.4 (89.9)	84.7 (89.3)	0.820	322
22	271 (346)	303 (355)	47 (61)	61.2(78.1)	66.0 (77.3)	0.647	438
				64.50	*64.11*	*0.7193*	*6,468*
				--------------- Weighted Averages --------------			*294*

Table 1d. Simulation Tests — Validity and Calibration Study — Results

	Col. 24	***Col. 25***	***Col. 26***	***Col. 27***	***Col. 28***	***Col. 29***
	Number of Items Correct Corresponding to a Score of 70% Correct		Number of Students Obtaining a Score of 70% Correct in This Sample		Percent of Students Obtaining a Score of 70% Correct in This Sample	
	Virginia SOL Test	TfHS Simulation Test	Virginia SOL Test	TfHS Simulation Test	Virginia SOL Test %	TfHS Simulation Test %
1	32	25	181	86	57.8	27.7
2	30	36	82	10	58.5	9.4
3	29	32	81	41	43.8	22.2
4	35	2	277	297	63.2	66.9
5	35	42	69	36	43.7	22.2
6	42	49	68	7	36.2	3.7
7	28	35	263	184	56.2	39.5
8	28	35	116	44	40.6	15.2
9	35	42	70	11	43.8	6.8
10	28	35	171	123	40.9	30.1
11	27	35	120	108	41.7	37.1
12	35	42	46	33	28.7	21.3
13	30	37	117	64	67.6	38.1
14	35	42	226	132	55.1	31.9
15	35	42	347	203	65.5	38.7
16	32	39	361	183	71.6	37.3
17	35	42	52	5	57.1	5.5
18	35	42	110	76	93.2	65.5
19	35	42	89	66	41.2	30.6
20	43	50	343	360	63.8	66.7
21	45	52	173	250	53.1	76.5
22	43	50	271	126	61.2	27.5
			3,633	*2,445*		

Table 2. ***TfHS Simulation Tests*** **Suggested Score Action Ranges**

Table with the ***original*** *cutting points for History & SS*	If the student's score is in this range, concentrate on teaching the precursor and underlying skills.	If the student's score is in this range, concentrate on teaching the SOL content in which the student appears weakest.	If the student's score is in this range, concen-trate on broadening, generalizing, and enrichment activites.
English, R & W, Grade K-3	**0 — 14**	**15 — 25**	**26 — 35**
English, Read. Lit., & Res., Grade 4-5	**0 — 14**	**15 — 31**	**32 — 52**
English, Read. Lit., & Res., Grade 6-8	**0 — 17**	**18 — 33**	**34 — 52**
Mathematics, Grade K-3	**0 — 35**	**36 — 48**	**49 — 60**
Mathematics, Grade 4-5	**0 — 26**	**27 — 43**	**44 — 60**
Mathematics, Grade 6-8	**0 — 20**	**21 — 36**	**37 — 70**
Science, Grade, K-3	**0 — 21**	**22 — 37**	**38 — 50**
Science, Grade, 4-5	**0 — 14**	**15 — 34**	**35 — 50**
Science, Grade, 6-8	**0 — 20**	**21 — 33**	**34 — 60**
History and Social Science, Grade K-3	**0 — 16**	**17 — 38**	**39 — 50**
History and Social Science, Grade 4-5	**0 — 21**	**20 — 40**	**41 — 50**
History and Social Science, Grade 6-8	**0 — 26**	**27 — 47**	**48 — 60**
English, Read. Lit., & Res., High School	**0 — 18**	**19 — 30**	**31 — 52**
Algebra I, High School	**0 — 13**	**14 — 40**	**41 — 60**
Algebra II, High School	**0 — 20**	**21 — 34**	**35 — 60**
Geometry, High School	**0 — 15**	**16 — 32**	**33 — 55**
Earth Science, High School	**0 — 13**	**14 — 32**	**33 — 60**
Biology, High School	**0 — 18**	**19 — 36**	**37 — 60**
Chemistry, High School	**0 — 19**	**20 — 39**	**40 — 60**
World History I, High School	**0 — 15**	**16 — 44**	**45 — 71**
World History II, High School	**0 — 26**	**27 — 59**	**60 — 73**
U. S. History, High School	**0 — 27**	**28 — 48**	**49 — 71**

Table 2a. *TfHS Simulation Tests* Suggested Score Action Ranges

*Table with the **NEW** cutting points for History & SS*	If the student's score is in this range, concentrate on teaching the precursor and underlying skills.	If the student's score is in this range, concentrate on teaching the SOL content in which the student appears weakest.	If the student's score is in this range, concen- trate on broadening, generalizing, and enrichment activites.
History and Social Science, Grade 4-5	**0 — 18**	**19 — 39**	**40 — 50**
History and Social Science, Grade 6-8	**0 — 18**	**19 — 42**	**43 — 60**
World History II, High School	**0 — 22**	**23 — 56**	**57 — 73**
U. S. History, High School	**0 — 25**	**26 — 41**	**42 — 71**

Table 3a. *TfHS* Grade Level Tests Suggested Score Action Ranges

Note: This table in an Extrapolition of the Actual Data	If the student's score is in this range, concentrate on teaching the precursor and underlying skills.	If the student's score is in this range, concentrate on teaching the SOL content in which the student appears weakest.	If the student's score is in this range, concentrate on broadening, generalizing, and enrichment activites.
English, R & W, Grade K	**0 — 13**	**14 — 23**	**24 — 32**
English, R & W, Grade 1	**0 — 11**	**12 — 20**	**21 — 28**
English, Read. Lit., & Res., Grade 2	**0 — 14**	**15 — 24**	**25 — 34**
English, Read. Lit., & Res., Grade 3	**0 — 15**	**16 — 26**	**27 — 37**
English, Read. Lit., & Res., Grade 4	**0 — 12**	**13 — 27**	**28 — 46**
English, Read. Lit., & Res., Grade 5	**0 — 13**	**14 — 30**	**31 — 50**
English, Read. Lit., & Res., Grade 6	**0 — 16**	**17 — 32**	**33 — 50**
English, Read. Lit., & Res., Grade 7	**0 — 15**	**16 — 30**	**31 — 47**
English, Read. Lit., & Res., Grade 8	**0 — 16**	**17 — 31**	**32 — 49**
English, Read. Lit., & Res., Grade 9	**0 — 17**	**18 — 34**	**35 — 53**
English, Read. Lit., & Res., Grade 10	**0 — 15**	**16 — 26**	**27 — 45**
English, Read. Lit., & Res., Grade 11	**0 — 17**	**18 — 29**	**30 — 50**
Mathematics, Grade K	**0 — 18**	**19 — 25**	**26 — 31**
Mathematics, Grade 1	**0 — 21**	**22 — 29**	**30 — 36**
Mathematics, Grade 2	**0 — 49**	**50 — 67**	**68 — 84**
Mathematics, Grade 3	**0 — 46**	**47 — 63**	**64 — 79**
Mathematics, Grade 4	**0 — 33**	**34 — 54**	**55 — 76**
Mathematics, Grade 5	**0 — 31**	**32 — 52**	**53 — 72**
Mathematics, Grade 6	**0 — 25**	**26 — 45**	**46 — 88**
Mathematics, Grade 7	**0 — 25**	**26 — 45**	**46 — 88**
Mathematics, Grade 8	**0 — 20**	**21 — 35**	**36 — 69**
Mathematics, Algebra I	**0 — 20**	**21 — 59**	**60 — 88**
Mathematics, Algebra II	**0 — 29**	**30 — 49**	**50 — 87**
Mathematics, Geometry	**0 — 23**	**24 — 48**	**49 — 82**

Table 3b. *TfHS* Grade Level Tests Suggested Score Action Ranges

Table with the OLDcutting points for History & SS ***Note: This table in an Extrapolition of the Actual Data***	If the student's score is in this range, concentrate on teaching the precursor and underlying skills.	If the student's score is in this range, concentrate on teaching the SOL content in which the student appears weakest.	If the student's score is in this range, concentrate on broadening, generalizing, and enrichment activites.
Science, Grade, K	**0 — 12**	**13 — 23**	**24 — 30**
Science, Grade, 1	**0 — 14**	**15 — 26**	**27 — 35**
Science, Grade, 2	**0 — 26**	**27 — 46**	**47 — 62**
Science, Grade, 3	**0 — 32**	**33 — 57**	**58 — 76**
Science, Grade, 4	**0 — 36**	**37 — 62**	**63 — 83**
Science, Grade, 5	**0 — 25**	**26 — 45**	**46 — 60**
Science, Grade, 6	**0 — 28**	**29 — 49**	**50 — 66**
Science, Grade, 7	**0 — 30**	**31 — 52**	**53 — 70**
Science, Grade, 8	**0 — 24**	**25 — 43**	**44 — 57**
Earth Science, High School	**0 — 41**	**42 — 71**	**72 — 95**
Biology, High School	**0 — 42**	**43 — 73**	**74 — 98**
Chemistry, High School	**0 — 42**	**43 — 73**	**74 — 98**
History and Social Science, Grade K	**0 — 12**	**13 — 23**	**24 — 30**
History and Social Science, Grade 1	**0 — 17**	**18 — 30**	**31 — 40**
History and Social Science, Grade 2	**0 — 33**	**34 — 58**	**59 — 78**
History and Social Science, Grade 3	**0 — 24**	**25 — 43**	**44 — 57**
History and Social Science, Grade 4	**0 — 21**	**22 — 38**	**39 — 50**
History and Social Science, Grade 5	**0 — 26**	**27 — 46**	**47 — 62**
History and Social Science, Grade 6	**0 — 29**	**30 — 51**	**52 — 68**
History and Social Science, Grade 7	**0 — 33**	**34 — 57**	**58 — 77**
World History I, High School	**0 — 39**	**40 — 67**	**68 — 90**
World History II, High School	**0 — 50**	**51 — 86**	**87 — 116**
U. S. History, High School	**0 — 36**	**37 — 63**	**64 — 84**

Table 3c. *TfHS* Grade Level Tests Suggested Score Action Ranges

Table with the NEWcutting points for History & SS ***Note: This table in an Extrapolition of the Actual Data***	If the student's score is in this range, concentrate on teaching the precursor and underlying skills.	If the student's score is in this range, concentrate on teaching the SOL content in which the student appears weakest.	If the student's score is in this range, concentrate on broadening, generalizing, and enrichment activites.
Science, Grade, K	**0 — 12**	**13 — 23**	**24 — 30**
Science, Grade, 1	**0 — 14**	**15 — 26**	**27 — 35**
Science, Grade, 2	**0 — 26**	**27 — 46**	**47 — 62**
Science, Grade, 3	**0 — 32**	**33 — 57**	**58 — 76**
Science, Grade, 4	**0 — 36**	**37 — 62**	**63 — 83**
Science, Grade, 5	**0 — 25**	**26 — 45**	**46 — 60**
Science, Grade, 6	**0 — 28**	**29 — 49**	**50 — 66**
Science, Grade, 7	**0 — 30**	**31 — 52**	**53 — 70**
Science, Grade, 8	**0 — 24**	**25 — 43**	**44 — 57**
Earth Science, High School	**0 — 41**	**42 — 71**	**72 — 95**
Biology, High School	**0 — 42**	**43 — 73**	**74 — 98**
Chemistry, High School	**0 — 42**	**43 — 73**	**74 — 98**
History and Social Science, Grade K	**0 — 12**	**13 — 23**	**24 — 30**
History and Social Science, Grade 1	**0 — 17**	**18 — 30**	**31 — 40**
History and Social Science, Grade 2	**0 — 33**	**34 — 58**	**59 — 78**
History and Social Science, Grade 3	**0 — 24**	**25 — 43**	**44 — 57**
History and Social Science, Grade 4	**0 — 19**	**20 — 37**	**38 — 50**
History and Social Science, Grade 5	**0 — 24**	**25 — 45**	**46 — 62**
History and Social Science, Grade 6	**0 — 22**	**23 — 46**	**47 — 68**
History and Social Science, Grade 7	**0 — 26**	**27 — 52**	**53 — 77**
World History I, High School	**0 — 39**	**40 — 67**	**68 — 90**
World History II, High School	**0 — 46**	**47 — 83**	**84 — 116**
U. S. History, High School	**0 — 34**	**35 — 56**	**57 — 84**

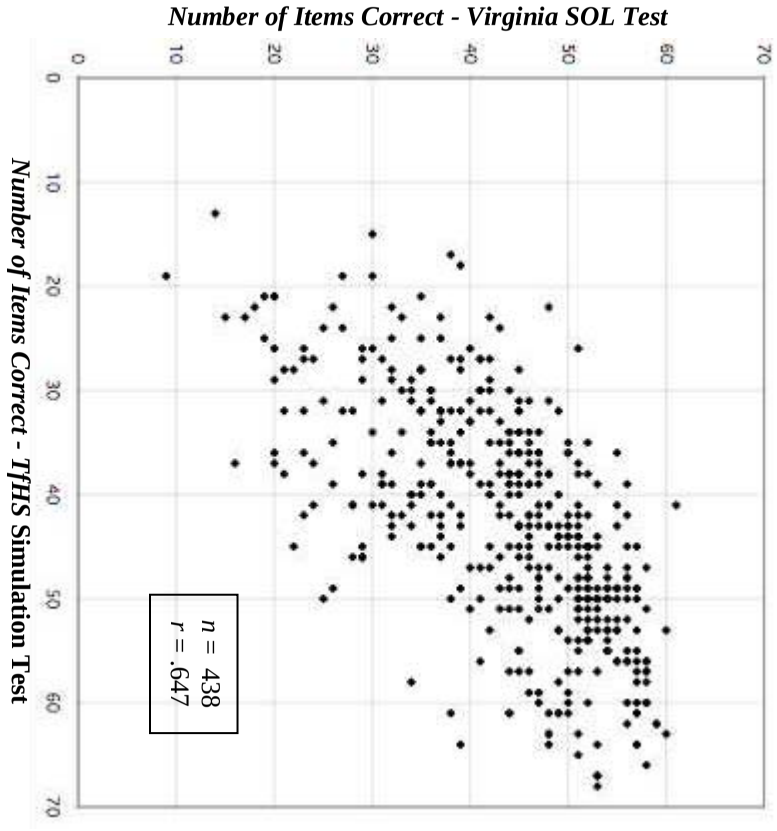


Figure 1 — US History

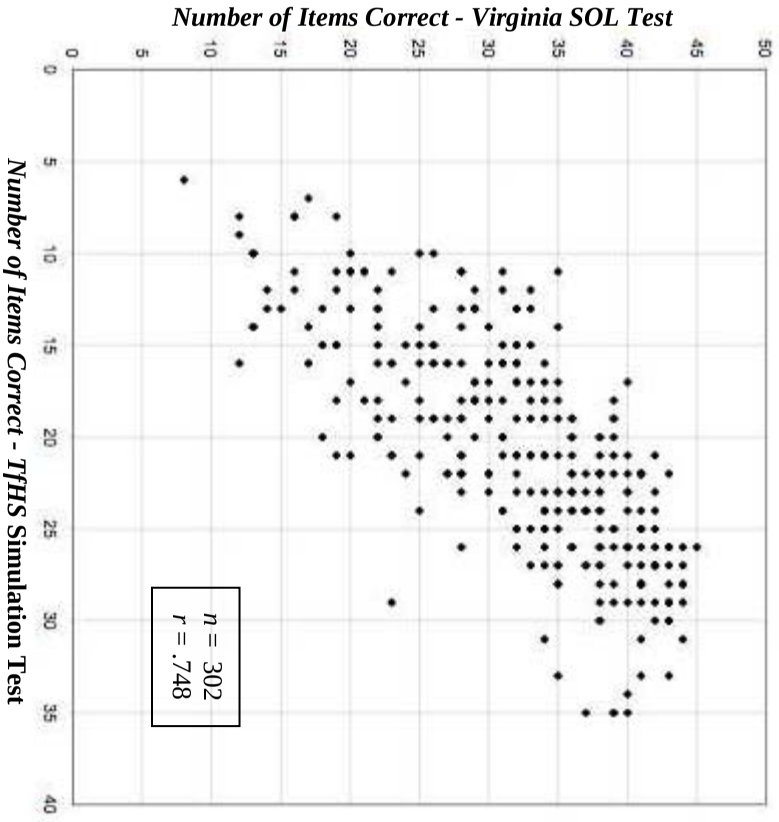


Figure 2 — Reading – Grade 3

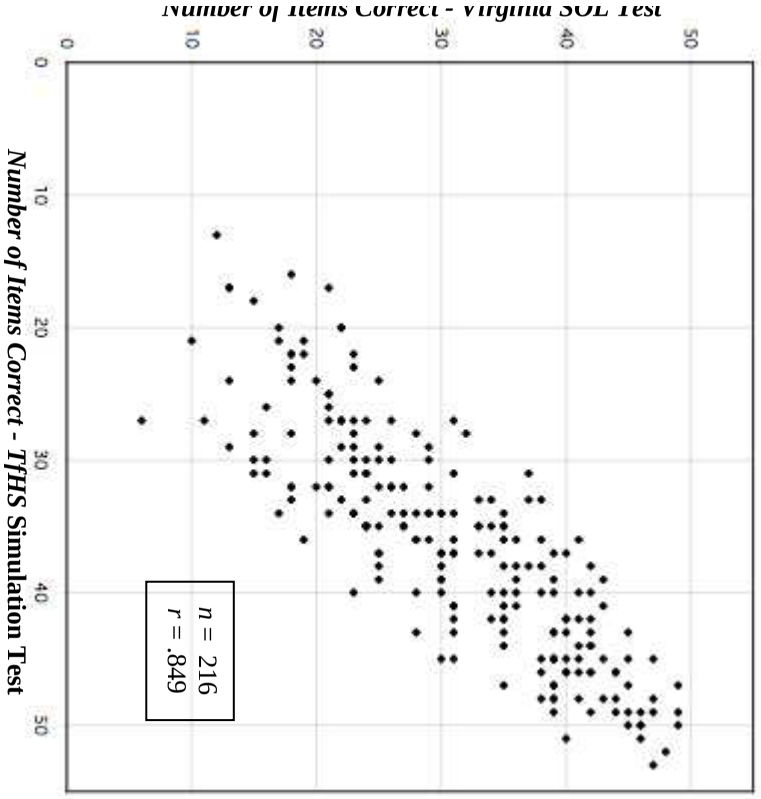


Figure 3 — Chemistry

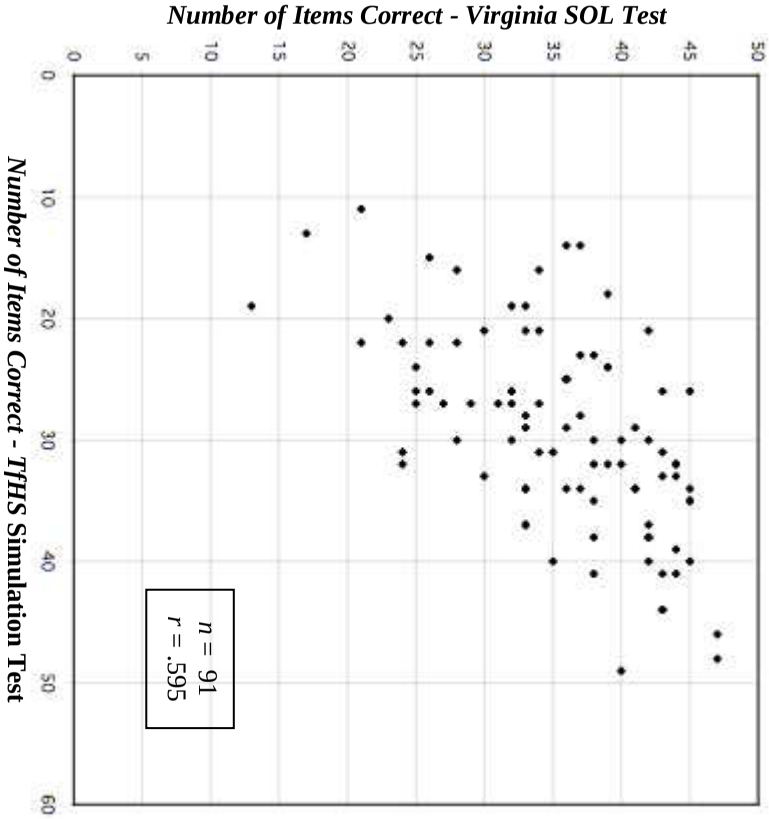


Figure 4 — Earth Science

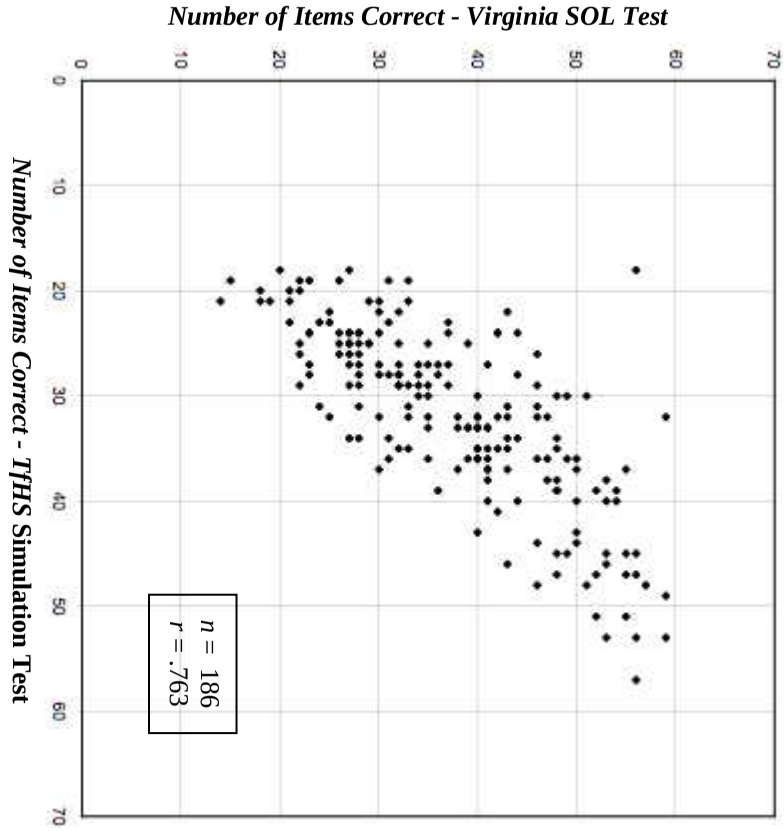


Figure 5 — Mathematics – Grade 8

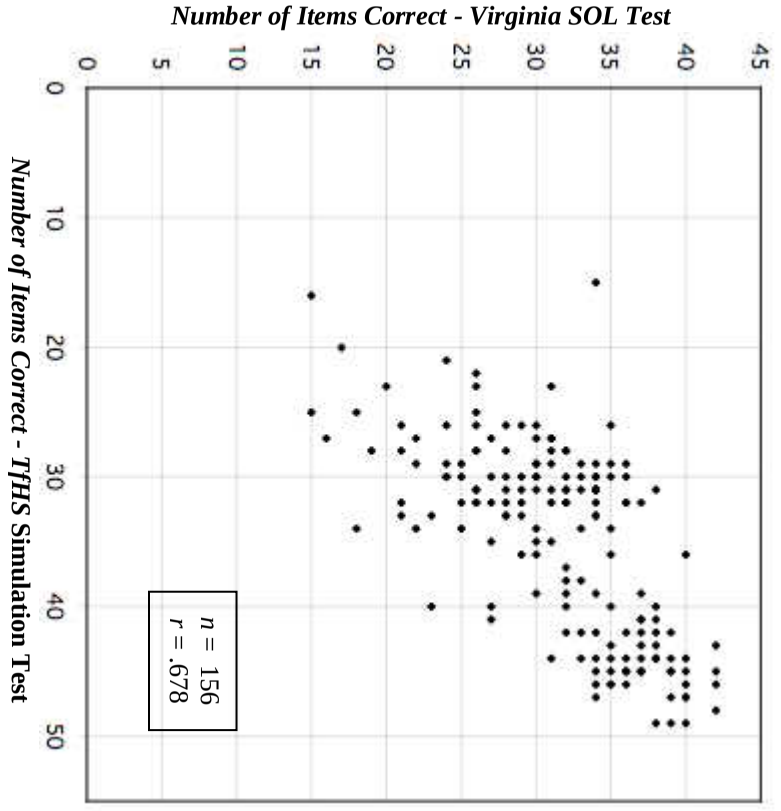


Figure 6 — Reading – End-of-Course