Listening. Learning. Leading.

A Concurrent Validity Study of the 2008 HSTW *Assessment Scores*

John W. Young
Fred Cline
Center for Validity Research
Educational Testing Service
Princeton, New Jersey 08541

December 2009

Author's Contact Information:
John W. Young, Ph.D.
Senior Research Scientist
& Research Director
Center for Validity Research
Educational Testing Service
MS 10-R
Princeton, New Jersey 08541
(609) 734-5037
jwyoung@ets.org

A Concurrent Validity Study of the 2008 HSTW *Assessment Scores*

Introduction

High Schools That Work (HSTW) is a school improvement initiative that was inaugurated by the Southern Regional Education Board (SREB) in 1987. At present, more than 1,200 *HSTW* sites in 32 states are using the framework of *HSTW* Goals and Key Practices to raise student achievement. To assess progress in school improvement and student achievement, one key component of *HSTW* is the *HSTW* Assessment, consisting of three subject tests (Mathematics, Reading, and Science). Because a new version of the assessment was developed and administered for the first time in 2008, it is important to conduct validity studies on the assessment results to evaluate whether the assessment is functioning as designed.

A New Test Design

In order to meet the goal of introducing entirely new assessments in all three subject areas for 2008, new *HSTW* frameworks in Mathematics, Reading, and Science were developed (Educational Testing Service, 2008). In previous years, the *HSTW* assessments were guided by the content of the 12th grade National Assessment of Educational Progress (NAEP) assessments in these three subject areas. The availability of new NAEP frameworks for 2009 for grade 12 in the three subject areas proved to be fortuitous timing for the development of the new *HSTW* assessments. The NAEP frameworks provided convenient starting points for tailoring and modifying those frameworks to align with the specific needs of *HSTW* --- primarily to measure readiness for college or the workplace; assess comprehension, analysis, and reasoning skills; and, for the first time, to provide individual student score reports on the *HSTW* assessments. The new *HSTW* assessments for 2008 were designed to meet a subset of the total set of skills and objectives as described in the corresponding NAEP frameworks. This subset of skills was determined by an external panel of

experts and was comprised of the skills and objectives deemed to be most critical in meeting the needs of *HSTW*. The recommendations of the external panel also reflected thinking that is consistent with professional organizations in education that have produced publications on the curriculum and assessment content in Mathematics, Reading, and Science.

This Study

The main purpose of a concurrent validity study is to evaluate one or more measures by investigating their relationship to other commonly used and established measures given at or about the same time to a group of individuals. For example, one way to judge the validity of scores from a new assessment is to see how well those scores correlate with current school grades. If the new assessment is designed to measure a student's knowledge or skills in a particular content area, we would expect that the scores from the assessment would correlate more highly with grades in the same content area and correlate less well with grades from other content areas. If the pattern of results conforms to our model of students' cognitive skills, then this provides one type of evidence that our interpretation of the test scores is valid and is supported by these results. A concurrent validity study differs from a predictive validity study, which attempts to show that a measure is valid for predicting future performance; for example, using SAT or ACT scores of high school students to predict college grades.

This validity study was conducted in the fall of 2008 by the Educational Testing Service (ETS) and SREB. When the *HSTW* Assessment was administered in January and February of 2008, *HSTW* school sites were asked if they would be interested in participating in future research studies. All *HSTW* schools in each of the six states (Georgia, Kentucky, Ohio, South Carolina, Texas, and West Virginia) that expressed interest in research participation were contacted. These states were targeted for participation because each has a large number of *HSTW* sites. Since the analyses for this study used scores from state-specific content assessments, it was necessary to ensure that there would

be a sufficient number of students from each state to achieve valid and reliable results. A total of 146 schools in the six states were contacted; of these, 68 schools agree to participate, although only 51 schools provided the data requested. *HSTW* sites were asked to complete a spreadsheet with students' information on cumulative high school grade point average (GPA), state test results, and national admissions (ACT or SAT) test results. In total, data from nearly 2,600 students were analyzed for this study.

Correlation Results

A common statistic used to measure the degree of association between two variables is the Pearson correlation coefficient, which ranges from -1.00 to +1.00, and is usually referred to simply as a correlation. Pearson correlations only measure linear (or straight-line) association; there are other statistical measures of associations that are more appropriate when one is interested in assessing more complex relationships between two variables (for example, a quadratic relationship). Positive correlations indicate that high values on one variable tend to be associated with high values on the second variable, and also that low values on the two variables tend to occur together. Negative correlations mean that high values on one variable tend to be associated with low values on the other variable. Since correlations are symmetric — that is, the correlation of variable A with variable B is the same as the correlation of B with A — it does not matter which variable is labeled first or second. As a general guide, in the social sciences, correlations between .00 and .20 are considered to be small; .21 to .50 to be moderate; and .51 to 1.00 to be large. It is also important to recognize that correlations only measure the degree of association, which is a weaker relationship than a cause-and-effect relationship between two variables. Pearson correlation coefficients were computed to determine the degree of association among scores on the *HSTW* Assessment, cumulative high school GPA, scores on state high school content tests, and scores on national college admissions tests.

The results showed that scores from the *HSTW* Assessment are generally moderately to highly correlated with all of the other measures of student achievement (see Tables 1 – 5). These analyses were conducted state-by-state because the state tests differ, such that using a single sample may have produced misleading results. The patterns for the validity coefficients are as expected: With few exceptions, *HSTW* Assessment scores correlated more highly with state test scores in the same content area (e.g., *HSTW* Reading with state-specific reading) than with state test scores in a different content area or with high school GPA (the correlations across content areas are not shown). It should be noted that the results for Texas are somewhat atypical: *HSTW* Assessment scores had substantially lower correlations with high school GPA than was found in the other five states. This does not appear to indicate a problem with the *HSTW* Assessment in Texas as we also found similarly low correlations between the Texas state test scores and high school GPA.

In Tables 4 and 5 are shown the correlations between *HSTW* Assessment scores and scores from the two national college admissions tests ACT and SAT. In these analyses, we were able to use a single sample since the same tests were taken by all students. Note that the samples in these analyses are smaller than the total sample for this study, as some students did not take either admissions test. With one exception, all of the correlations between the HSTW Assessment scores and the corresponding section scores from the ACT or SAT are above .50, which can be characterized as large. The only exception was a correlation of .28 between *HSTW* Reading and SAT Writing, but because these represent different skill areas, one would expect the correlation to be moderate. It is important to note that because the admissions tests are high-stakes tests for high school students, one would expect that students would be highly motivated to perform well. Given that assumption, the fact that the *HSTW* Assessment scores are so highly correlated with those scores provides evidence that these students also appeared to be motivated when they took the *HSTW* Assessment. In other words, if a number of students were not motivated to do well on the *HSTW* Assessment, the correlations with the admissions tests would have been much lower.

Summary

The results from these analyses provide strong empirical support for the concurrent validity of the new *HSTW* Assessment, given for the first time in 2008. The results showed that the scores from the *HSTW* Assessment are generally moderately to highly correlated with all of the other measures of student achievement, and that the correlations fit an expected pattern since the *HSTW* Assessment scores correlated more highly with state test scores in the same content area than with state test scores in a different content area or with high school GPA. In addition, all of the correlations between the *HSTW* Assessment scores and the corresponding section scores from both of the national college admissions tests are substantial, which appears to indicate that low student motivation was not an issue for the *HSTW* Assessment, at least for the subset of students in this study who also took the ACT or SAT.

Reference

Educational Testing Service. (2008). *High Schools That Work technical manual*. Princeton, NJ: Author.

Tables

Table 1

Correlations of HSTW *Reading with Other Measures of Student Achievement*

	HS GPA		State Reading		State Writing	
State	Correlation	N	Correlation	N	Correlation	N
GA	0.42	349	0.41	338	0.29	332
KY	0.56	447	0.66	116	0.47	214
OH	0.45	560	0.56	538	0.48	539
SC	0.55	674	0.64	557	—	—
TX	0.26	133	0.52	291	—	—
WV	0.57	310	0.56	193	—	—

Note: South Carolina, Texas, and West Virginia do not include a separate writing section in their state-wide test.

Table 2

Correlations of HSTW *Mathematics with Other Measures of Student Achievement*

	HS GPA		State Mathematics		State Science	
State	Correlation	N	Correlation	N	Correlation	N
GA	0.41	349	0.48	338	0.38	337
KY	0.52	447	0.70	270	0.59	270
OH	0.48	560	0.50	538	0.55	538
SC	0.43	674	0.64	557	—	—
TX	0.25	133	0.55	289	0.56	290
WV	0.58	310	0.67	192	0.57	192

Note: South Carolina does not include a separate science section in their state-wide test.

Table 3

Correlations of HSTW *Science with Other Measures of Student Achievement*

	HS GPA		State Science		State Mathematics	
State	Correlation	N	Correlation	N	Correlation	N
GA	0.45	349	0.47	337	0.40	338
KY	0.39	447	0.71	270	0.66	270
OH	0.38	560	0.55	538	0.39	538
SC	0.44	674	—	—	0.57	557
TX	0.17	133	0.49	290	0.48	289
WV	0.53	310	0.60	192	0.51	192

Note: South Carolina does not include a separate science section in their state-wide test.

Table 4

Correlations of HSTW *Assessment with SAT*

	SAT Reading		SAT Writing		SAT Mathematics	
	Correlation	N	Correlation	N	Correlation	N
HSTW Reading	0.54	790	0.28	689	—	—
HSTW Mathematics	—	—	—	—	0.59	734
HSTW Science	—	—	—	—	0.56	734

Note: Correlations across subject areas were not computed as they are not meaningful.

Table 5

Correlations of HSTW *Assessment with ACT*

	ACT Reading		ACT English		ACT Mathematics		ACT Science	
	Correlation	N	Correlation	N	Correlation	N	Correlation	N
HSTW Reading	0.57	1127	0.59	1127	—	—	—	—
HSTW Mathematics	—	—	—	—	0.67	1118	0.62	1098
HSTW Science	—	—	—	—	0.55	1118	0.56	1098

Note: Correlations across subject areas were not computed as they are not meaningful.