



Research Memorandum

ETS RM-14-06

Recommending Passing Scores for the High School Equivalency Test (*HiSET*®)

Richard J. Tannenbaum

Clyde M. Reese

September 2014

ETS Research Memorandum Series

EIGNOR EXECUTIVE EDITOR

James Carlson
Principal Psychometrician

ASSOCIATE EDITORS

Beata Beigman Klebanov
Research Scientist

Heather Buzick
Research Scientist

Brent Bridgeman
Distinguished Presidential Appointee

Keelan Evanini
Managing Research Scientist

Marna Golub-Smith
Principal Psychometrician

Shelby Haberman
Distinguished Presidential Appointee

Donald Powers
Managing Principal Research Scientist

Gautam Puhan
Senior Psychometrician

John Sabatini
Managing Principal Research Scientist

Matthias von Davier
Director, Research

Rebecca Zwick
Distinguished Presidential Appointee

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ayleen Stellhorn
Editor

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Memorandum series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Memorandum series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

Recommending Passing Scores for the High School Equivalency Test (*HiSET*[®])

Richard J. Tannenbaum and Clyde M. Reese
Educational Testing Service, Princeton, NJ

September 2014

Find other ETS-published reports by searching the ETS ReSEARCHER
database at <http://search.ets.org/researcher/>

To obtain a copy of an ETS research report, please visit
<http://www.ets.org/research/contact.html>

Action Editor: Gautam Puhan

Reviewers: Priya Kannan and Caroline Wylie

Copyright © 2014 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, HiSET, and LISTENING. LEARNING. LEADING.
are registered trademarks of Educational Testing Service (ETS).



Abstract

Standard-setting workshops were conducted with 84 educators to recommend passing scores for the 5 *HiSET*[®] tests: Language Arts–Reading, Language Arts–Writing, Mathematics, Science, and Social Studies. All tests included multiple-choice items; the HiSET Language Arts–Writing test also included an essay. A variation of a modified Angoff method was applied to the multiple-choice items, and a variation of an Extended Angoff method was applied to the essay prompt. When expressed as a percentage of the number of multiple-choice points available, the passing scores ranged from 52% to 60%; the median was 58%. The recommended passing score for the essay prompt was 3 on a 6-point rubric. Procedural and internal sources of evidence support the reasonableness of the recommended passing scores.

Key words: HiSET, standard setting, cut scores, passing scores

Recently released statistics from the U.S. Department of Education show high school graduation rates at a 40-year high; approximately 80% of students in 2012 graduated from high school (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2014). While we should applaud the steady increase in the graduation rate, we should not lose sight of the flipside of that statistic—approximately one in five students still do not graduate—and minority students, economically disadvantaged students, English-language learners, and students with disabilities had graduation rates lower than the national average. In 2013, the unemployment rate for people 25 years old and older was 11% for those with less than a high school diploma compared to 7.5% for those who had graduated high school but with no college credits (U.S. Department of Labor, Bureau of Labor Statistics, 2013). Unemployment rates are even lower for those with some college education and those with bachelor's and other advanced degrees.

The *HiSET*[®] program is intended to offer individuals who did not complete a traditional high school experience an opportunity to obtain a high school equivalency certificate. The HiSET Program includes five tests, each measuring one specific subject-matter area: Language Arts–Reading, Language Arts–Writing, Mathematics, Science, and Social Studies.

A series of standard-setting workshops were conducted in April 2014 to recommend passing scores for each of the five HiSET tests. This report documents the standard-setting procedures and results of those workshops.

Standard Setting

Setting a standard for a test means establishing the minimum score that differentiates two levels of proficiency (Tannenbaum, 2011). The minimum score, often referred to as a passing score, delineates the boundary between one level of proficiency and the next higher level of proficiency. The passing score is the lowest acceptable score that signifies entrance into the higher level. For example, in K-12 student testing, a passing score may differentiate students performing at a *proficient* level from students performing at a *basic* level. When test scores are used to classify individuals into different levels or categories, the process of standard setting is a critical contributor to the cumulative evidence supporting the validity of the test score interpretation and use (Bejar, Braun, & Tannenbaum, 2007; Kane, 2006; Margolis & Clauser, 2014; Tannenbaum & Kannan, in press).

Setting standards has long been recognized to be a judgment-based process of, in effect, forming a policy or decision rule (Cizek, 1993; Kane, 2001). There are no empirically correct passing scores (O’Neill, Buckendahl, Plake, & Taylor, 2007). The concept of how much knowledge or skill must be demonstrated on a test, and embodied by a test score, to reach a level of proficiency or performance, is a function of the values and expectations of those involved in setting the standard (O’Neill et al., 2007; Tannenbaum & Katz, 2013). In this value-based context, an evaluation of the credibility and meaningfulness of the passing score—the reasonableness of the passing score—is based on the appropriateness of the standard-setting design and the quality of the implementation of the standard-setting process (Tannenbaum & Cho, 2014).

Overview of the HiSET Test Battery

Each of the five tests constituting the HiSET program includes multiple-choice items; the HiSET Language Arts–Writing test also includes an essay prompt. For the HiSET Language Arts–Reading, Mathematics, Science, and Social Studies tests, the total raw score for each test is the number of multiple-choice items answered correctly. The total raw score is then scaled so it is reported on the HiSET 20-point scale. The process of scaling supports the common meaning of scores across different forms of the same test. A similar process is used for the HiSET Language Arts–Writing test, where the raw scores for the multiple-choice and essay portions of the test are separately scaled and then added to produce the reported HiSET Writing score. The specific HiSET tests are described below. More detailed information about each test is available from the following website: <http://hiset.ets.org/about/content>.

Language Arts–Reading

The HiSET Language Arts–Reading test consists of 40 multiple-choice items presented as unique sets of items (e.g., seven to nine items) associated with different reading passages. A test taker has 65 minutes to complete the test. Text selections include both literary and informational texts. The items measure four process categories—(a) comprehension, (b) inference and interpretation, (c) analysis, and (d) synthesis and generalization.

Language Arts–Writing

The HiSET Language Arts–Writing test consists of 50 multiple-choice items and one essay prompt. A test taker has 75 minutes to complete the multiple-choice portion and 45 minutes to complete the essay. The essay is scored using a 6-point rubric. The multiple-choice section covers three categories: (a) organization of ideas, (b) language facility, and (c) writing conventions. The essay covers these same three categories, but also a fourth category, (d) development of ideas.

Mathematics

The HiSET Mathematics test consists of 50 multiple-choice items. A test taker has 90 minutes to complete the test. The test covers both content and processes. The four major content categories are (a) numbers and operations on numbers; (b) measurement and geometry; (c) data analysis, probability, and statistics; and (d) algebraic concepts. The process categories are (a) understand mathematical concepts and procedures, (b) analyze and interpret information, and (c) synthesize data and solve problems.

Science

The HiSET Science test consists of 50 multiple-choice items. A test taker has 80 minutes to complete the test. The test covers both content and processes. The three major content categories are (a) life science, (b) physical science, and (c) earth science. The process categories are (a) interpret and apply, (b) analyze, and (c) evaluate and generalize.

Social Studies

The HiSET Social Studies test consists of 50 multiple-choice items. A test taker has 70 minutes to complete the test. The test covers both content and processes. The four major content categories are (a) history, (b) civics and government, (c) economics, and (d) geography. The process categories are (a) interpret and apply, (b) analyze, and (c) evaluate and generalize.

Method

Panelists

The outcomes of standard setting are directly associated with score interpretation and use; therefore, standard setting becomes an integral part of the overall validity argument (Bejar et al., 2007; Kane, 2006). As a judgment-based process that relies on the opinions of experts, the

selection of qualified experts is critical to the meaningfulness of the resulting passing scores (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Tannenbaum & Katz, 2013).

The issuance of a HiSET certificate is intended to provide some assurance that the holder has demonstrated a level of proficiency in core academic subject areas comparable to that expected of a high school graduate (Educational Testing Service, 2013). In this context, two fundamental perspectives are essential to be represented by the experts (educators) assembled for the standard-setting workshops: one is the perspective of educators who work with students seeking a more traditional high school diploma, and the other is the perspective of educators who work more directly with students (often including adults) seeking a high school equivalency certificate. This dual representation helps assure that the performance standard for each test is aligned with the intended meaning of the HiSET certificate.

Table 1 presents a summary of the educators' demographic information. Four separate panels were assembled; one of the panels focused on the two language arts tests: HiSET Language Arts–Reading and HiSET Language Arts–Writing. The panels, collectively, included 84 educators representing 26 states and Washington, DC; no educator served on more than one panel. Approximately half of the educators (43) represented traditional high school settings, and approximately half (41) represented adult education settings. Most of the educators were current teachers or instructional providers (75 of the 84 educators).

Language Arts–Reading and Writing

One panel of 19 educators recommended passing scores for the two HiSET Language Arts tests. Ten educators were high school language arts teachers, and nine worked in adult education settings (five taught in an adult education program, one taught in a correctional facility, two were administrators for adult education programs, and one was an adult education curriculum developer). The educators represented 13 states: Kentucky, Louisiana, Maine, Missouri, Montana, North Carolina, New Hampshire, New Jersey, New York, Nevada, Tennessee, Wisconsin, and West Virginia.

Table 1***Educator Demographics (Across Panels)***

	<i>N</i>	%
Current position		
High school teacher	40	48%
Adult education provider/instructor	28	33%
Correctional education provider/instructor	7	8%
Administrator (high school setting)	3	4%
Administrator (adult education setting)	5	6%
Adult education curriculum developer	1	1%
Race		
White	66	79%
Black or African American	14	17%
Hispanic or Latino	2	2%
Asian or Asian American	1	1%
Native Hawaiian or Other Pacific Islander	1	1%
Gender		
Female	59	70%
Male	25	30%
Years of experience as an educator ^a		
3 years or less	5	6%
4–7 years	30	36%
8–11 years	15	18%
12–15 years	12	14%
16 years or more	21	25%

^aOne educator did not provide years of experience.

Mathematics

Twenty-two educators recommended a passing score for the HiSET Mathematics test. Thirteen educators worked in a high school setting (12 were high school mathematics teachers, and one was an administrator), and nine worked in adult education settings (seven taught in an

adult education program, one taught in a correctional facility, and one was an administrator). The educators represented 14 states: Arkansas, California, Idaho, Louisiana, Maine, Mississippi, Missouri, Montana, North Carolina, Nebraska, New Hampshire, New Jersey, Nevada, and Rhode Island.

Science

Twenty-two educators recommended a passing score for the HiSET Science test. Nine educators worked in a high school setting (eight were high school science teachers, and one was an administrator), and 13 worked in adult education settings (eight taught in an adult education program, three taught in a correctional facility, and two were administrators). The educators represented 15 states: Idaho, Kentucky, Louisiana, Maine, Missouri, Mississippi, North Carolina, New Hampshire, New Jersey, Ohio, South Carolina, Tennessee, Texas, Wisconsin, and Wyoming.

Social Studies

Twenty-one educators recommended a passing score for the HiSET Social Studies test. Eleven educators worked in a high school setting (10 were high school social studies teachers, and one was an administrator), and 10 worked in adult education settings (eight taught in an adult education program, and two taught in a correctional facility). The educators represented 14 states and the District of Columbia: Arkansas, California, Kentucky, Louisiana, Maryland, Maine, Missouri, Montana, North Dakota, New Hampshire, New Jersey, South Dakota, Tennessee, and Vermont.

Procedures

The selection of an appropriate standard-setting method is informed by the structure of the test (types of test questions or tasks), by the evidence available supporting the quality of the method, and by the experience of the standard-setting facilitator with the selected method (Hambleton & Pitoniak, 2006; Tannenbaum & Cho, 2014). Taking into consideration these criteria, two standard-setting methods were applied. A variation of a modified Angoff method was applied to the multiple-choice items, and an Extended Angoff method was applied to the essay prompt of the writing test (Plake & Cizek, 2012). Both methods share some common features, and these features are similarly shared across most widely used standard-setting

methods (Hambleton, Pitoniak, & Coppella, 2012; Tannenbaum & Katz, 2013). The major steps followed to implement these approaches are described below.

Reviewing the Test

The standard-setting panelists (educators) completed the test and self-scored. This was followed by a discussion focusing on the major content areas being addressed by the test. In particular, the educators were asked to remark on what was expected of test takers seeking a high school equivalency certificate, what content areas might be particularly challenging for these test takers, and what content areas were most important for test takers to understand. These familiarization activities helped bring the educators to a shared understanding of what the test was and was not covering; provided them with a better understanding of the difficulty of the tested content; and helped to bring to light any preconceived expectations held by the educators, which might have introduced *construct-irrelevant variance* later in the standard-setting process.

Defining the Performance Standard

The educators then described the minimally sufficient knowledge and skills expected of a test taker who should be classified as passing the test. This is the performance standard, referred to here as the just qualified candidate (JQC), and serves as the frame of reference against which standard-setting judgments are made (Tannenbaum & Katz, 2013). The focus of the standard-setting task is to identify the test score that, on average, best aligns with this performance expectation. The emphasis on minimally sufficient knowledge and skills when describing the JQC is purposeful. This is because the passing score, which is the numeric equivalent of the performance expectations described in the JQC on the score scale, is intended to be the lowest acceptable score that denotes entrance into the passing category. The educators drew upon their experience with having reviewed the test, and their own teaching experiences with high school students or students seeking a high school equivalency certificate, and they constructed the JQC description through both independent small-group work and then whole-panel consensus.

Educators' Judgments

Multiple-choice items. A variation of a probability-based modified Angoff method (Brandon, 2004; Tannenbaum & Katz, 2013) was applied to the multiple-choice items. Each educator judged each item on the likelihood (probability or chance) that the JQC would answer

the item correctly. Educators made their judgments using a rating scale that ranged from 0 to 1, in increments of .05. The lower the value, the less likely it is that the JQC would answer the item correctly, because the item is difficult for the JQC. The higher the value, the more likely it is that the JQC would answer the item correctly.

The educators were instructed to approach this task in two stages: firstly, to locate the difficulty of the item for the JQC into one of three bands: 0 to .35 (*difficult*), .40 to .65 (*moderately difficult/easy*), and .70 to 1 (*easy*); secondly, to decide on the single probability value within the band. For example, if, to be answered correctly, an item required knowledge that was clearly aligned with the knowledge expected of the JQC, then this was likely an *easy* item for the JQC, and so it would first be located in the .70 to 1 band. The second decision would be to identify the specific probability value within that band (.70, .75, .80, .85, .90, .95, or 1).

Essay prompt. A variation of Extended Angoff method (Hambleton & Plake, 1995; Tannenbaum & Katz, 2013) was applied to writing test essay prompt. For this portion of the study, educators decided on the rubric score value that would most likely be earned by the JQC. They were instructed to review the description of JQC then review the essay prompt and scoring rubric; each of the 6 points on the rubric was anchored by a sample response (benchmark).

Training and practice. Before engaging in the formal standard-setting task, the educators were trained in how to carry out the task and provided an opportunity to practice making their standard-setting judgments. This included a discussion of their judgment rationales. The opportunity for educators to practice and discuss their judgments helps to clarify and reinforce the reasoning they will need to engage in as they make their standard-setting judgments (Tannenbaum & Cho, 2014). At the conclusion of this set of activities, educators completed an evaluation form addressing the adequacy of the training and their readiness to proceed. All educators (on each panel) confirmed their understanding of the standard-setting task and readiness to proceed.

Multiple rounds of judgment. Following a first round of independent judgments, feedback was provided to the panel. One level of feedback included each educator's recommended passing score (unidentified), in addition to the mean, median, highest and lowest passing scores, and the standard deviation. This level of feedback occurred both for the multiple-choice items and for the one essay prompt. A second level of feedback was included for the multiple-choice items. For each item, the distribution of the educators' probability judgments

across the three bands of difficulty (0 to .35, .40 to .65, .70 to 1) was presented, along with the mean of the educators' judgments, and the empirical difficulty of the item. The empirical difficulty (p -value) represents the percentage of a national sample of more than 20,000 high school seniors who answered the item correctly. Although high school seniors are not the target test-taking population, their responses nonetheless provide useful comparative information for the educators to consider as they reflect on their own item judgments.

Discussion of the feedback focused on the reasonableness of the educators' mean passing-score recommendation and on the rationales for the item-specific judgments. During the item discussion, educators were asked to share their decision-making process, making explicit the connection between what the item required of a test taker and the knowledge and skills expected of a JQC. The educators were informed that the p -values provided another source of information to consider but that the values should *not* be taken as the *correct* values or the values they should move their judgments toward. The educators were instructed to consider when their mean item judgment was very different from that of the p -value and simply to consider that information, along with the discussion of judgment rationales, to decide if they wanted to make a change to their Round 1 item judgments during the second round of judgments.

For the essay prompt, the opportunity to change their Round 1 judgments occurred after discussion of the panel-level feedback (mean, median, highest and lowest passing scores, and the standard deviation). The results of the Round 2 judgments for each test were shared with the educators at the end of the study.

Final evaluations. The educators completed an evaluation form at the conclusion of their study addressing the quality of the standard-setting implementation and their acceptance of the recommended passing score. The responses to the evaluation provide evidence of the validity of the standard-setting process and the reasonableness of the passing score (Hambleton & Pitoniak, 2006; Kane, 2001).

Results

The results for each HiSET test are presented next. Results are provided for each round of judgment. The means of the Round 2 judgments, following discussion of the Round 1 results, are the passing-score recommendations.

Language Arts–Reading

Table 2 summarizes the standard-setting judgments of the educators for the HiSET Language Arts–Reading test. The table shows the passing scores (the sum of the item-level judgments) recommended by each educator. The test includes 40 multiple-choice items, and the passing-score recommendations are the educators’ judgments of the number of items that must be answered correctly to be considered performing at the level of a JQC. The panel’s recommended passing score, the mean of the recommendations of the 19 educators at the end of Round 2, was 23 out of 40. The computed mean score was 22.67. This value was rounded to the next highest whole score of 23. The rule to round to the next highest whole score reflects that, on average, the panel recommended a raw passing score greater than 22. This general policy of rounding to the next highest whole number was followed for all the final passing-score recommendations. The mean and median were consistent between Rounds 1 and 2. The standard deviation of recommendations across the educators decreased from Round 1 to Round 2, as often occurs following discussion.

Language Arts–Writing

Table 3 summarizes the standard-setting judgments of the educators for the multiple-choice portion of the HiSET Language Arts–Writing test. The test includes 50 multiple-choice items. The panel’s recommended passing score was 28 out of 50 (the mean score was rounded to the next highest whole score). The mean decreased from 29 to 28 points between Rounds 1 and 2; the median was consistent. The standard deviation of recommendations across the educators decreased from Round 1 to Round 2.

Given that there was only one essay prompt, a table of results is not presented. The Round 1 distribution of recommendations yielded 14 recommendations of a score of 3 and five recommendations of a score of 4. The Round 2 distribution yielded 18 recommendations of a score of 3 and only one of a score of 4. The median of the Round 2 judgments, 3 on the 6-point rubric, is the recommended passing score.

Table 2***Passing-Score Recommendation—Reading***

Educator		
ID	Round 1	Round 2
1	21.45	22.20
2	22.00	22.40
3	23.90	24.10
4	21.70	22.15
5	27.50	21.50
6	20.25	21.60
7	18.05	20.35
8	21.20	21.65
9	25.80	25.70
10	24.35	24.05
11	22.95	21.95
12	25.85	25.95
13	23.80	22.85
14	25.25	25.00
15	22.20	21.20
16	26.80	24.95
17	25.15	24.95
18	18.60	20.55
19	15.80	17.55
Mean	22.77	22.67
Median	22.95	22.20
<i>SD</i>	3.12	2.14
SEJ	.71	.49
SEM		2.9

Note. SEJ = standard error of judgment; SEM = standard error of measurement.

Table 3***Passing-Score Recommendation—Writing Multiple-Choice Items***

Educator		
ID	Round 1	Round 2
1	31.25	29.80
2	22.65	24.70
3	28.15	28.10
4	29.00	29.35
5	26.50	26.50
6	26.70	27.15
7	26.85	28.45
8	27.85	27.20
9	39.85	37.65
10	28.60	28.90
11	28.10	26.80
12	36.05	30.20
13	22.15	23.45
14	29.35	28.40
15	26.40	25.55
16	31.55	31.05
17	32.75	31.60
18	21.50	23.00
19	18.70	20.65
Mean	28.10	27.82
Median	28.10	28.10
<i>SD</i>	5.02	3.71
SEJ	1.15	.85
SEM		3.6

Note. SEJ = standard error of judgment; SEM = standard error of measurement.

Mathematics

Table 4 summarizes the standard-setting judgments of the educators for the HiSET Mathematics test. The test includes 50 multiple-choice items. The panel's recommended passing score was 30 out of 50 (the mean score was rounded to the next highest whole score). The mean was consistent between Rounds 1 and 2; the median (rounded to the next highest whole number) decreased from 31 to 29 points. The standard deviation of recommendations across the educators decreased from Round 1 to Round 2.

Science

Table 5 summarizes the standard-setting judgments of the educators for the HiSET Science test. The test includes 50 multiple-choice items. The panel's recommended passing score was 26 out of 50 (the mean score was rounded to the next highest whole score). The mean decreased from 28 to 26 points between Rounds 1 and 2, and the median (rounded to the next highest whole number) decreased from 28 to 27 points. The standard deviation of recommendations across the educators increased slightly from Round 1 to Round 2.

Social Studies

Table 6 summarizes the standard-setting judgments of the educators for the HiSET Social Studies test. The test includes 50 multiple-choice items. The panel's recommended passing score was 30 out of 50 (the mean score was rounded to the next highest whole score). The mean increased from 29 to 30 points between Rounds 1 and 2; the median was consistent. The standard deviation of recommendations across the educators decreased from Round 1 to Round 2.

Table 4***Passing-Score Recommendation—Mathematics***

Educator		
ID	Round 1	Round 2
1	25.20	25.90
2	26.20	26.15
3	33.15	32.50
4	33.95	33.50
5	27.95	27.90
6	32.90	33.05
7	32.60	33.65
8	28.25	28.50
9	29.00	28.80
10	33.20	31.35
11	26.20	26.00
12	27.00	28.95
13	30.50	31.05
14	32.35	33.15
15	30.15	30.40
16	32.35	31.15
17	25.85	28.30
18	29.95	28.10
19	30.40	26.45
20	28.10	27.25
21	31.25	28.55
22	33.50	33.55
Mean	30.00	29.74
Median	30.28	28.88
<i>SD</i>	2.81	2.69
SEJ	.60	.57
SEM		3.4

Note. SEJ = standard error of judgment; SEM = standard error of measurement.

Table 5***Passing-Score Recommendation—Science***

Educator		
ID	Round 1	Round 2
1	24.90	24.10
2	28.65	27.25
3	30.10	30.05
4	38.05	33.90
5	23.65	21.00
6	33.15	30.00
7	27.05	25.25
8	26.70	25.35
9	26.90	25.90
10	17.45	15.00
11	30.25	30.05
12	27.55	26.15
13	30.85	29.05
14	22.30	16.25
15	21.95	21.80
16	32.70	31.05
17	24.70	25.60
18	28.45	28.05
19	30.20	28.30
20	26.30	26.45
21	22.35	22.25
22	27.30	25.70
Mean	27.34	25.84
Median	27.18	26.03
<i>SD</i>	4.48	4.55
SEJ	.96	.97
SEM		3.4

Note. SEJ = standard error of judgment; SEM = standard error of measurement.

Table 6***Passing-Score Recommendation—Social Studies***

Educator		
ID	Round 1	Round 2
1	28.25	29.10
2	30.25	31.60
3	25.85	27.05
4	32.10	31.95
5	30.60	30.70
6	32.85	31.50
7	36.55	35.30
8	23.70	24.95
9	24.95	24.70
10	26.35	25.40
11	31.60	31.75
12	25.20	24.65
13	35.20	35.05
14	31.05	31.05
15	26.95	27.30
16	31.00	30.50
17	23.90	25.35
18	19.50	20.75
19	30.80	30.35
20	31.55	31.05
21	29.85	30.20
Mean	28.95	29.06
Median	30.25	30.35
<i>SD</i>	4.16	3.72
SEJ	0.91	0.81
SEM		3.4

Note. SEJ = standard error of judgment; SEM = standard error of measurement.

Sources of Evidence Supporting the Passing Scores

Kane (1994, 2001) noted that the reasonableness of and confidence placed on passing scores are strengthened by evidence of the quality of the standard-setting implementation (procedural evidence), and by evidence of the consistency of the educators' judgments and the likelihood of replicating the recommended passing scores (internal evidence). Results addressing these two sources of evidence are presented below.

Procedural evidence. The educators' responses to the training and end-of-workshop evaluations are important sources of procedural evidence (Cizek, 2012; Cizek & Bunch, 2007). Noted previously, all 84 educators verified at the conclusion of the training and practice opportunity that they understood the purpose of setting standards, understood the steps they were to follow to complete the standard-setting task, and confirmed their readiness to begin their first round of independent judgments.

Further, following the completion of each workshop, the participating educators were asked (a) if they understood the purpose of the workshop, (b) if the provided instructions and explanations were clear, and (c) if they were adequately trained. Across the four panels, all 84 educators agreed or strongly agreed that they understood the purpose; 83 of 84 agreed or strongly agreed that the instructions and explanations were clear; and all 84 agreed or strongly agreed that they were adequately trained. The educators also provided feedback about the general implementation of the workshops. Between 81 and 84 educators reported that the workshops were efficient, coordinated, fair, understandable, and satisfying.

In addition to the educators' evaluation of the standard-setting process, they also were shown their panel's recommended passing score (the panel mean following Round 2) and asked (a) how comfortable they were with the recommended passing score and (b) if they thought the score was too high, too low, or about right. Table 7 summarizes their judgments of the recommended passing scores. Across the HiSET tests, the majority of educators indicated they were very comfortable with the passing score; the median was 68.5%, and it ranged from 59% for science to 91% for mathematics. The majority of educators also judged the recommended passing scores to be about right; the median was 89.5%, and it ranged from 82% for science to 100% for writing (essay). The collective results from the evaluations—both the judgments of the standard-setting process and the recommended passing scores—support the quality of the standard-setting implementation.

Table 7***Evaluation of Recommended Passing Scores***

	Comfort with the recommended passing score							
	Very comfortable		Somewhat comfortable		Somewhat uncomfortable		Very uncomfortable	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
	<hr/>							
Reading	12	63	7	37	0	0	0	0
Writing (MC)	12	63	7	37	0	0	0	0
Writing (Essay)	14	74	5	26	0	0	0	0
Mathematics	20	91	1	5	0	0	1	5
Science	13	59	8	36	1	5	0	0
Social Studies	17	81	3	14	1	5	0	0
	Appropriateness of the recommended passing score							
	Too low		About right		Too high			
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%		
	<hr/>							
Reading	1	5	16	84	2	11		
Writing (MC)	1	5	16	84	2	11		
Writing (Essay)	0	0	19	100	0	0		
Mathematics	0	0	21	95	1	5		
Science	2	9	18	82	2	9		
Social Studies	0	0	20	95	1	5		

Note. MC = multiple choice.

Internal evidence. Two aspects of internal evidence (consistency) were documented. The first is the correlation of the educators' average item judgments for the multiple-choice items with the empirical *p*-values, and the second is the likelihood of replicating the recommended passing scores. The first criterion addresses the ability of the educators to recognize the relative difficulty of the test items; the higher the correlation, the stronger the evidence of the educators' ability to discern relative difficulty. If the educators cannot reasonably recognize the *relative* difficulty of the items, that would call into question the meaningfulness of their *absolute* item judgments, which is the standard-setting task (Tannenbaum & Cho, 2014). The correlations

ranged from .45 (for the writing and social studies tests) to .83 (for the mathematics test); the median correlation across the five tests was .78. These results indicate that the educators were reasonably able to differentiate the relative difficulty of the multiple-choice items. Moderate correlations between educators' average item judgments and *p*-values are not uncommon (e.g., Brandon, 2004; Tannenbaum & Kannan, in press).

The second aspect of internal evidence addresses the likelihood of replicating the recommended passing scores. One way to evaluate this is by conducting the standard setting with two different panels of educators and comparing the two recommended passing scores (see Tannenbaum & Kannan, in press). However, when only one panel has been formed, an approximation of replicability is provided by the standard error associated with the recommended passing scores (Cizek & Bunch, 2007, Kaftandjieva, 2010). This standard error of judgment (SEJ) is an index of the extent to which the passing score would vary if the workshop were repeated with different panels of educators (Zieky, Perie, & Livingston, 2008). The smaller the value, the less likely it is that other panels would recommend a significantly different passing score. A general guideline for interpreting the SEJ is its size relative to the standard error of measurement (SEM) of the test. According to Cohen, Kane, and Crooks (1999), an SEJ less than one-half of the SEM is considered reasonable. The SEJs and SEMs are reported in Tables 2–6. In all instances, the SEJ was well below half the value of the SEM, ranging from .17 for the reading and mathematics tests to .29 for the science test; the median across the tests was .24. These collective results (correlations and SEJs) support the consistency of the educators' judgments and recommended passing scores.

Summary and Next Steps

The goal of the HiSET program is to provide individuals with an opportunity to obtain a high school equivalency diploma. In this score-use context, passing scores are needed to differentiate between individuals who have and have not demonstrated sufficient knowledge and skills to earn the equivalency diploma. Standard-setting workshops were conducted with 84 educators representing both traditional high school settings and adult education settings to recommend passing scores for each of the five HiSET tests: Language Arts–Reading, Language Arts–Writing, Mathematics, Science, and Social Studies. Final passing score recommendations for each test are presented in Table 8. When expressed as a percentage of the maximum number of points available on a test, the passing scores for the multiple-choice tests ranged from 52% for

the science test to 60% for the mathematics and social studies tests; the median across the multiple-choice tests was 58%. The recommended passing score of 3 for the essay prompt reflects a limited response, according to the rubric. (A score of 2 reflects a weak response, and a score of 4, an adequate response.)

Table 8

Recommended Passing Scores by Test

Test	Recommended passing score
Language Arts–Reading (40 points maximum)	23
Language Arts–Writing (multiple-choice items, 50 points maximum)	28
Language Arts–Writing (essay prompt, 6 points maximum)	3
Mathematics (50 points maximum)	30
Science (50 points maximum)	26
Social Studies (50 points maximum)	30

Although both procedural and internal sources of evidence support the reasonableness of the recommended passing scores, such evidence does not mean that the recommended passing scores are somehow immune to reconsideration and adjustment. Setting a standard is comparable to forming a policy, where decisions are neither right nor wrong (Kane, 2001). In fact, it is common for decision makers (those ultimately responsible for setting the operational passing scores) to adjust the recommended passing scores (Geisinger & McCormick, 2010). In this regard, a governing board (state-level decision makers) will be convened to review the results of the standard-setting workshops, along with additional data (e.g., passing rate outcomes, external benchmarks of college readiness) to inform its evaluation of the recommended passing scores and any suggested passing-score adjustments.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Bejar, I. I., Braun, H. I., & Tannenbaum, R. J. (2007). A prospective, progressive, and predictive approach to standard setting. In R. W. Lissitz (Ed.), *Assessing and Modeling Cognitive Development in School* (pp. 1–30). Maple Grove, MN: JAM Press.
- Brandon, P. R. (2004). Conclusions about frequently studied modified Angoff standard-setting topics. *Applied Measurement in Education, 17*, 59–88.
- Cizek, G. J. (1993). Reconsidering standards and criteria. *Journal of Educational Measurement, 30*, 93–106.
- Cizek, G. J. (2012). The forms and functions of evaluations in the standard setting process. In G. J. Cizek (Ed.), *Setting performance standards: foundations, methods, and innovations* (2nd ed., pp. 163–178). New York, NY: Routledge.
- Cizek, G. J. & Bunch, M. (2007). *Standard setting: A practitioner's guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Cohen, A. S., Kane, M. T., & Crooks, T. J. (1999). A generalized examinee-centered method for setting standards on achievement tests. *Applied Measurement in Education, 12*, 343–366.
- Educational Testing Service. (2013). *HiSET™ information brief*. Retrieved from <http://hiset.ets.org/s/pdf/taag.pdf>
- Geisinger, K. F., & McCormick, C. M. (2010). Adopting cut scores: Post-standard-setting panel considerations for decision makers. *Educational Measurement: Issues and Practice, 29*, 38–44.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. *Educational Measurement, 4*, 433–470.
- Hambleton, R. K, Pitoniak, M. J., & Coppella, J. M. (2012). Essential steps in setting performance standards. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 47–76). New York, NY: Routledge.
- Hambleton, R. K., & Plake, B. S. (1995). Using an Extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education, 8*, 41–56.

- Kaftandjieva, F. (2010). *Methods for setting cut scores in criterion-referenced achievement tests: A comparative analysis of six recent methods with an application to tests of reading in EFL*. Arnhem, The Netherlands: CITO.
- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, *64*, 425–462.
- Kane, M. T. (2001). So much remains the same: Conceptions and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 53–88). Mahwah, NJ: Erlbaum.
- Kane, M. T. (2006). Validation. *Educational Measurement*, *4*, 17–64.
- Margolis, M. J., & Clauser, B. E. (2014). The impact of examinee performance information on judges' cut scores in modified Angoff standard-setting exercises. *Educational Measurement: Issues and Practice*, *33*, 15–22.
- O'Neill, T. R., Buckendahl, C. W., Plake, B. S., & Taylor, L. (2007). Recommending a nursing-specific passing standard for the IELTS examination. *Language Assessment Quarterly*, *4*, 295–317.
- Plake, B. S., & Cizek, G. J. (2012). Variations on a theme: The modified Angoff, Extended Angoff, and yes/no standard setting methods. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 181–199). New York, NY: Routledge.
- Tannenbaum, R. J. (2011). Standard setting. In J. W. Collins & N. P. O'Brien (Eds.), *Greenwood Dictionary of Education* (2nd ed.). Santa Barbara, CA: ABC-CLIO.
- Tannenbaum, R. J., & Cho, Y. (2014). Critical factors to consider in evaluating standard-setting studies to map language test scores to frameworks of language proficiency. *Language Assessment Quarterly*, *11*, 233–249.
- Tannenbaum, R. J., & Kannan, P. (in press). Consistency of Angoff-based judgments: Are item judgments and passing scores replicable across different panels of experts? *Educational Assessment*.
- Tannenbaum, R. J., & Katz, I. R. (2013). Standard setting. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Vol 3. Testing and assessment in school psychology and education* (pp. 455–477). Washington, DC: American Psychological Association.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2014). *Public high school four-year on-time graduation rates and event dropout rates: School years 2010–11 and 2011–12*. Retrieved from <http://nces.ed.gov/pubs2014/2014391.pdf>.

U.S. Department of Labor, Bureau of Labor Statistics (2013). Labor force statistics from the current population survey. Retrieved from <http://www.bls.gov/cps/cpsaat07.pdf>.

Zieky, M. J., Perie, M., & Livingston, S. A. (2008). *Cutscores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.