Using PISA to Internationally Benchmark State Performance Standards

Gary W. Phillips &
Tao Jiang

American Institutes for Research
April 8, 2011

Paper presentation at the annual conference of the American Educational Research Association, New Orleans, Louisiana 2011.

Using PISA to Internationally Benchmark State Performance Standards

Gary W. Phillips & Tao Jiang[1]

Abstract

This study describes how the Programme for International Student Assessment (PISA) was used for internationally benchmarking state performance standards. The process is accomplished in two steps. First, PISA items are embedded in the administration of the state assessment and calibrated on the state scale. The international item calibrations are then used to link the state scale to the PISA scale through common item linking. The second step is to use the statistical linking as part of the state standard setting process to help standard setting panelists determine how high their state standards need to be in order to be internationally competitive. This process was carried out in Delaware, Hawaii, and Oregon, and results are reported here for two of the states: Hawaii and Delaware.

Key words: Equating, linking, item response theory, international benchmarking.

Introduction

In 2010, the American Institutes for Research obtained permission from the Organization for Economic Cooperation and Development (OECD) to use secure items from the Programme for International Student Assessment (PISA) for purposes of linking state assessments within the United States to the PISA scale. The OECD provided a representative sample of 30 secure PISA items in Reading, Mathematics, and Science. The PISA items covered the 2006 and 2009 PISA assessment cycles. In addition to the PISA items, the Australian Council for Educational Research (ACER), which is the current vender for the OECD contracted to conduct PISA, provided the international item parameters and their standard errors, as well as the linear transformations needed to link the state assessments to the PISA scale. The administration, security, and scoring of the PISA items were carried out by the American Institutes for Research

[1] Phillips, G. W., & Jiang, T. (2011). *Using PISA to internationally benchmark state performance standards*. Paper presented at the Annual Conference of the American Educational Research Association, New Orleans, LA.

(AIR) based on a License Agreement between AIR and the OECD and monitored by the National Center for Education Statistics (NCES).

Method

In each of the three states, the Bookmark method of standard setting was used (Mitzel, Lewis, Patz, & Green, 2001). In the Bookmark approach, an ordered-item booklet is assembled representing the content standards and test blueprint used in the state for operational testing. The items in the ordered-item booklet are ordered from easiest (in the front of the booklet) to the hardest (at the end of the booklet). A broadly representative panel of stakeholders (teachers, educators, community leaders, business leaders, and parents) is then asked to use the ordered-item booklet to recommend performance standards on the test. In addition to the ordered-item booklet, the panelists use the state content standards (statements about what students should know and be able to do) and performance level descriptors (statements about how much students should know and be able to do) to recommend standards. The panelists are trained to locate the item in the ordered-item booklet that meets a certain response probability. For example, if 100 students were at the Proficient cut-score, which item would 67% of the students answer correctly? Standards are set over several rounds, facilitating a consensus among the panelists. The international benchmarking described in the paper was introduced to the panelists as a way for them to benchmark their recommended performance standards.

PISA was used for the international benchmarking. PISA is sponsored by the OECD, which is headquartered in Paris, France. PISA is administered to 15-year-old students every three years and covers science, mathematics, and reading literacy. In each cycle, a different subject is emphasized. In 2006 (which was the most recent PISA data at the time of this study) the emphasis was science. The 2006 U.S. PISA sample consisted of students who were in grades 8 (1%), 9 (11%), 10 (71%), and 11 (17%). Therefore, the modal grade of the 2006 U.S. PISA sample was grade 10. For this reason, the international benchmarking was done in grade 10 in each of the states.

Statistical linking of the state test to PISA was done as follows. Suppose there is a set of P items from PISA (representative of the PISA content domain) that are common between the state test (X) and PISA (Y). The international items are administered on both the state test and PISA, which represent the administration of the common items at two different times and two different samples of examinees. Statistical linking involves the use of these common items to

rescale the state test so that it is on the PISA scale. Since PISA uses the Rasch model, the discussion that follows assumes the item response theory (IRT) model used in common item linking is the one-parameter Rasch model (and the one-parameter partial credit model for constructed response items). Using the Rasch Model, let the common items be indexed by $i_1,\ldots,i_P$. The two sets of Rasch difficulty parameters are $d_1^Y,\ldots,d_P^Y$ and $d_1^X,\ldots,d_P^X$. For a partial credit item (Masters, 1982), the above notation represents the difficulty of the partial credit item, which is the average of the parameters of the item. The linking constant based on mean/mean equating is

$$\hat{\delta}_L = \frac{\sum_{k=1}^{P} d_k^X}{P} - \frac{\sum_{k=1}^{P} d_k^Y}{P}, \tag{1}$$

and the standard error of linking can be estimated by

$$\hat{\sigma}_L = \frac{1}{P}\sqrt{\sum_{k=1}^{P} se^2\left(d_k^X\right) + \sum_{k=1}^{P} se^2\left(d_k^Y\right)}. \tag{2}$$

Once the standard error of linking is determined, it is used to determine the standard error of the PISA-equivalent performance standards:

$$\hat{\sigma}_z = \sqrt{\hat{\sigma}_{RP}^2 + \hat{\sigma}_L^2}. \tag{3}$$

The statistic $\hat{\sigma}_{RP}^2$ is the error variance of the response probability of the items from the state test used in the ordered-item booklet described below for PISA. (The derivation of this statistic is in the Appendix. The derivation for the three-parameter IRT model can be found in Phillips and Jiang, 2011).

The PISA items were used in three statesð Delaware, Hawaii, and Oregonð to help inform the standard setting process. Each state used the Bookmark method of standard setting and wanted to have international benchmarks as part of its standard setting process. The 30 items were embedded as field test items in the computer adaptive test in 2010 in each state. The computer adaptive algorithm randomly administered the items just like it would any other field

test item. The items were randomly administered across the state in the same manner as all other field test items. For example, letôs say the state had $F = 400$ items for field testing, $P = 30$ PISA items, and the test length was $L = 40$, then the expected number of PISA items administered to each student was $E(p) = 1/F * L * P = 3$. If the student population size was $N = 10{,}000$, then the expected sample size per PISA item was $E(n) = 1/F * L * N = 1{,}000$. The number of items per student does not provide enough information to get estimates of PISA scores (i.e., distributions of PISA-plausible values) but it does provide enough students to get good state-based estimates of the PISA item parameters. This then permits the state test to be linked to the PISA scale through common item linking.

Results

If the state item parameters and the PISA item parameters are measuring similar constructs, then we would expect the state item parameter estimates to be correlated with the international PISA item parameter estimates. The correlations are presented in Table 1. In general, the correlations are very high, especially in mathematics and reading.

Table 1: Correlation Between State and International Item Parameters

	Delaware	Hawaii	Oregon
Science	0.75	0.78	0.83
Math	0.94	0.96	0.96
Reading	0.89	0.91	0.93

The common PISA linking items were calibrated on the state scale, and then the Rasch linking constants and standard errors of linking were obtained. In some cases, one item was omitted because enough students did not respond to some score categories, which prevented AIR from using the PISA rubric for scoring. The number of common items, the linking constants, and the standard error of the linking constants are presented in Tables 2ï 4.

Table 2: Number of Common Items, Linking Constant and Standard Error of Linking for Delaware DCAS Linked to PISA

	Common Items	Linking Constant	Linking Error	Linking Error in PISA Metric
Science	30	-0.00801	0.01857	1.732003
Math	30	2.49361	0.02379	1.853071
Reading (males)	29	1.61536	0.02049	1.643095
Reading (females)	29	1.61536	0.02049	1.627452

Reading (missing gender)	29	1.61536	0.02049	1.644399

Table 3: Number of Common Items, Linking Constant and Standard Error of Linking for Hawaii HAS Linked to PISA

	Common Items	Linking Constant	Linking Error	Linking Error in PISA Metric
Science	29	0.04038	0.01784	1.6636685
Math	30	1.060387	0.02272	1.7697516
Reading (males)	29	1.179908	0.02234	1.7913274
Reading (females)	29	1.179908	0.02234	1.7742730
Reading (missing gender)	29	1.179908	0.02234	1.7927486

Table 4: Number of Common Items, Linking Constant and Standard Error of Linking for Oregon OAKS Linked to PISA

	Common Items	Linking Constant	Linking Error	Linking Error in PISA Metric
Science	30	3.87439	0.01259	1.1740355
Math	30	3.75395	0.01872	1.4584961
Reading (males)	29	3.20565	0.01249	1.0012539
Reading (females)	29	3.20565	0.01249	0.9917214
Reading (missing gender)	29	3.20565	0.01249	1.0020482

The linear transformations needed to place each state scale on to the PISA scale were provided by the OECD and are contained in Tables 5ï 7. Only the state-specific linking constants were added to the linear transformations. The linking error in the PISA metric varies in reading because PISA uses a different linear transformation in reading depending on gender and whether the data are missing (see Adams, 2009, page 246). Because of this, the PISA-equivalents and standard errors of PISA-equivalents for reading in Tables 10, 13, and 15 below were determined as weighted averages (weighted by sample sizes for males, females, and missing gender).

Table 5: Hawaii Transformations (puts Hawaii HSA on the PISA scale)

Science: ((Hawaii Science Theta - 0.0403805172413793- 0.1797)/1.0724) * 100 + 500
Math: ((Hawaii Mathematics Theta -1.060387+0.0752) + 0.1344) / 1.2838* 100 + 500
Reading Girls: ((0.8739 * (Hawaii Reading Theta - 1.1799075862069) + 0.0970 - 0.5076) / 1.1002) * 100 + 500 Boys: ((0.8823 * (Hawaii Reading Theta - 1.1799075862069) + 0.0204- 0.5076) / 1.1002) * 100 + 500 Missing gender: ((0.883* (Hawaii Reading Theta - 1.1799075862069) + 0.0552- 0.5076) / 1.1002)* 100 + 500

Table 6: Delaware Transformations (puts Delaware DCAS on the PISA scale)

Science: ((Delaware Science Theta - (-0.00800566666666668) - 0.1797)/1.0724) * 100 + 500
Math: ((Delaware Mathematics Theta -2.493607+0.0752) + 0.1344) / 1.2838* 100 + 500
Reading Girls: ((0.8739 * (Delaware Reading Theta - 1.61535517241379) + 0.0970 - 0.5076) / 1.1002) * 100 + 500 Boys: ((0.8823 * (Delaware Reading Theta - 1.61535517241379) + 0.0204- 0.5076) / 1.1002) * 100 + 500 Missing gender: ((0.883* (Delaware Reading Theta - 1.61535517241379) + 0.0552- 0.5076) / 1.1002) * 100 + 500

Table 7: Oregon Transformations (puts Oregon OAKS on the PISA Scale)

Science: ((Oregon Science Theta - 3.87439166666667- 0.1797)/1.0724) * 100 + 500
Math: (Oregon Mathematics Theta -3.75394766666667-0.0405+ 0.1344) / 1.2838* 100 + 500
Reading Girls: ((0.8739 * (Oregon Reading Theta - 3.20565206896552) + 0.0970 - 0.5076) / 1.1002) * 100 + 500 Boys: ((0.8823 * (Oregon Reading Theta - 3.20565206896552) + 0.0204- 0.5076) / 1.1002) * 100 + 500 Missing gender: ((0.883* (Oregon Reading Theta - 3.20565206896552) + 0.0552- 0.5076) / 1.1002) * 100 + 500

The results of the international benchmarking are reported in Tables 8ï 15 below. The international benchmarking in science for Oregon is not presented because the standard setting for science is not scheduled until spring 2011. Table 9 for Hawaii mathematics will be used for illustrative purposes. An ordered-item booklet with 61 page numbers was used to facilitate the standard setting in mathematics in Hawaii. Column 1 shows the page numbers in the ordered-item booklet that were used to set mathematics performance standards in the state. Column 2 shows the item parameters associated with the page numbers in the ordered-item booklet, and column 3 contains the thetas associated with the response probability for each item. Column 4 notes the standard error of the theta associated with the response probability. This standard error is needed because it is a component of error in the standard error of the PISA-equivalents.

Column 5 includes the state impact data used during the standard setting. This column tells the panelists what percent of students are likely to reach the performance standard depending on which page number is selected. Column 6 shows the PISA-equivalent of the theta in column 3. These were determined through the linear transformations provided in Tables 5ï 7. Column 7 is the standard error of the PISA-equivalent as noted in equation 3. Column 8 shows the inverse cumulative percent for the population of 30 OECD countries that participated in the 2006 PISA. It shows the approximate percentage of students across the OECD countries that would reach each PISA-equivalent performance standard. Finally, column 9 shows the inverse cumulative percent for the population of students in Finland that participated in the 2006 PISA. It shows the approximate percentage of students in Finland (the highest performing country in the 2006 PISA) that would reach each PISA-equivalent performance standard. Data from Tables 8ï 16 were used to construct PowerPoint slides that communicated the results of the international benchmarking to the standard setting panelists. Each state used the data in different ways.

The previous Meets Proficient performance standard for mathematics in Hawaii was on page 19 of the ordered-item booklet. The standard setting panelists decided to place the new Hawaii Meets Proficient performance standard on page 26. This raised the Hawaii Meets Proficient performance standard from a standard comparable in difficulty to Level II of PISA to one that is comparable in difficulty to Level III of PISA and similar to the international average of the OECD student performance.

In general, it is better to carry out this procedure when embedding field test items in the operational assessment rather than independent field tests: If students are less motivated in the field test, then they will be less motivated on the PISA itemsð resulting in an underestimate of PISA performance.

Inferential Limits of State-PISA Linking

One immediate reaction to embedding PISA items in state tests might be to cite a report issued by the National Research Council (Koretz, Bertenthal, and Green, 1999) at the National Academy of Sciences (NAS) that recommends against this practice. The main idea of the recommendations by the NAS was that embedding items from national assessments (or in our case, international assessments) ñwill not provide valid, reliable and comparable national [or in our case, international] scores for individual students.ò (Page 3) This paper agrees with the

conclusion in the NAS report related to scores for individual students. However, the report did not make any recommendations about the use of embedded items for estimating aggregate information. The report concludes, ñGiven the limited data available on this issue, the committee does not offer a conclusion about the use of embedding to obtain aggregate information.ò The current study uses embedding to do just thatð estimate the aggregate scale alignment between the state scale and the PISA scale. This procedure should not be used to estimate individual scores or used for subgroup disaggregating within the state. It should only be used for the purpose described here - to internationally benchmark the state achievement levels against the international scale. Embedding PISA items in the state assessment is an example of aggregate statistical linking (Thissen, 2007). The process does not yield PISA scores or PISA-equivalent scores for individual students. Instead, it yields a linking between the state scale and the PISA scale. Part of the licensing agreement between AIR and the OECD sets limits on inferences from the state-linking study. The licensing agreement states,

> *It is important to understand that embedding PISA items in state achievement tests will result in a PISA-equivalent scale, not an actual PISA scale. The purpose of this exercise is to obtain an international benchmark for the state achievement levels. If Licensee wants more information about PISA results (disaggregation, statistical relationships between PISA and background variables) then it should administer the full PISA assessment to a state representative sample of students. In addition, it should be noted that PISA does not provide scaled scores for individual students, so there will be no PISA-equivalent scale score for individual students. Licensor will provide a concordance table that indicates the PISA metric that is equivalent to the state scaled score metric. This will result in obtaining the PISA-equivalent of the state achievement standard. The term "PISA-equivalent" refers to the PISA score that is statistically equivalent in difficulty to each scaled score on the state test. This will be accomplished by Licensor using a common item linking procedure. Psychometrically, "linking' is distinguishable from "equating." In equating the two tests are comparable in content and reliability. In linking the two tests are similar (but not identical) in content and not equally reliable.*

Table 8: International Benchmarking With PISA for Hawaii, Science, 2010

Ordered-Item Booklet Page	Item Parameter	Theta Associated With RP67	Standard Error of Theta Associated With RP67	Percent Reaching State Cut-Score	PISA-Equivalent Scaled Score	Standard Error of PISA-Equivalent	Percent Reaching OECD PISA-Equivalent	Percent Reaching Finland PISA-Equivalent
65	2.28	3.0	0.18	1	760	17	0	1
PISA Level VI = 708 (Top Performers)								
64	1.10	1.8	0.11	2	648	10	6	16
63	0.95	1.7	0.10	3	634	9	8	21
PISA Level V = 633 (Top Performers)								
62	0.82	1.5	0.10	4	622	9	10	25
61	0.79	1.5	0.10	4	620	9	10	26
60	0.76	1.5	0.10	4	617	9	11	27
59	0.71	1.4	0.10	5	612	9	12	28
58	0.65	1.4	0.09	5	606	9	13	31
57	0.63	1.3	0.09	5	604	9	14	31
56	0.63	1.3	0.09	5	604	9	14	31
55	0.60	1.3	0.09	6	601	9	14	33
54	0.54	1.3	0.09	6	596	9	16	35
53	0.54	1.3	0.09	6	596	9	16	35
52	0.53	1.2	0.09	6	595	9	16	36
51	0.53	1.2	0.09	6	595	9	16	36
50	0.51	1.2	0.09	7	593	9	16	36
49	0.40	1.1	0.09	7	583	9	19	41
48	0.40	1.1	0.09	7	582	8	19	41
47	0.37	1.1	0.09	8	580	8	20	42
46	0.34	1.0	0.09	8	577	9	21	44
45	0.33	1.0	0.09	9	577	9	21	44
44	0.30	1.0	0.09	9	574	9	22	45
43	0.27	1.0	0.09	11	571	8	23	46
PISA Level IV = 559 (Strong Performers)								
42	-0.61	0.8	0.07	11	558	7	27	52
41	0.10	0.8	0.09	12	555	8	28	54
40	0.09	0.8	0.09	13	554	8	28	54
39	0.04	0.8	0.09	14	550	8	30	56
38	0.00	0.7	0.09	14	546	8	32	58
37	-0.05	0.7	0.09	15	541	8	33	60
36	-0.10	0.6	0.09	16	536	8	35	62
35	-0.11	0.6	0.09	18	535	8	36	63
34	-0.16	0.5	0.09	20	531	8	37	65
33	0.78	0.4	0.07	21	518	6	43	70
32	-0.32	0.4	0.08	23	515	8	44	71
31	-0.40	0.3	0.08	24	508	8	47	74
30	-0.74	0.3	0.07	25	507	7	47	74
29	-0.44	0.3	0.08	26	505	8	48	75
28	-0.50	0.2	0.08	26	499	8	51	77
27	-0.58	0.1	0.09	33	491	8	54	80

PISA Level III = 484 (Moderate Performers)								
26	-0.66	0.1	0.09	36	484	8	57	82
25	-0.66	0.0	0.09	38	484	8	57	82
24	-0.71	0.0	0.09	40	479	8	59	84
23	-0.72	0.0	0.09	43	478	8	59	84
22	-0.89	-0.2	0.09	46	463	8	65	88
21	-0.95	-0.2	0.09	50	457	8	68	89
20	-1.06	-0.3	0.09	53	447	8	71	91
19	-1.15	-0.4	0.09	56	438	8	74	93
18	-1.10	-0.5	0.09	59	437	8	75	93
17	-1.17	-0.5	0.09	59	437	9	75	93
16	-1.27	-0.6	0.09	63	427	9	78	94
15	-1.28	-0.6	0.09	63	426	9	78	94
14	-0.87	-0.6	0.07	64	426	7	78	94
13	-0.80	-0.6	0.06	66	425	6	78	95
12	-1.31	-0.6	0.09	67	423	9	79	95
11	-1.38	-0.7	0.09	70	417	9	81	96
PISA Level II = 409 (Moderate Performers)								
10	-1.48	-0.8	0.10	73	407	9	84	96
9	-1.50	-0.8	0.10	75	405	9	84	97
8	-1.52	-0.8	0.10	76	403	9	85	97
7	-1.65	-0.9	0.10	79	392	10	87	98
6	-1.92	-1.2	0.11	86	366	10	92	99
5	-1.95	-1.2	0.11	87	364	10	92	99
4	-2.09	-1.4	0.11	90	351	10	94	99
3	-1.71	-1.5	0.10	93	337	10	96	100
PISA Level I = 335 (Lowest Performers)								
2	-2.63	-1.9	0.14	97	300	13	98	100
1	-2.85	-2.1	0.15	98	280	14	99	100

Note: The 2006 PISA inverse cumulative percentages (percent reaching the PISA-equivalents) in science were not available to the authors. Therefore, the distributions were estimated assuming a normal approximation (30 OECD countries, mean = 500, standard deviation = 95; Finland, mean = 563, standard deviation = 86). Any country or combination of countries that participated in 2006 PISA can be selected for comparisons. The result of the standard setting was to raise the Meets Proficiency cut-score (this is the performance standard used for federal accountability reporting) from page 28 in the ordered-item booklet to page 30.

Table 9: International Benchmarking With PISA for Hawaii, Mathematics, 2010

Ordered-Item Booklet Page	Item Parameter	Theta Associated With RP67	Standard Error of Theta Associated With RP67	Percent Reaching State Cut-Score	PISA-Equivalent Scaled Score	Standard Error of PISA-Equivalent	Percent Reaching OECD PISA-Equivalent	Percent Reaching Finland PISA-Equivalent
61	1.67	2.5	0.12	6	627	10	8	17
60	1.56	2.3	0.11	7	610	9	11	22
59	1.54	2.3	0.11	7	609	9	11	23
PISA Level V = 607 (Top Performers)								
58	1.50	2.2	0.11	8	606	9	12	24
57	1.44	2.1	0.11	8	601	8	13	26
56	1.36	2.1	0.11	9	595	8	15	28
55	1.34	2.0	0.10	9	593	8	15	29
54	1.29	2.0	0.10	9	590	8	16	30
53	1.25	2.0	0.10	10	586	8	17	32
52	1.22	1.9	0.10	10	584	8	18	33
51	1.21	1.9	0.10	10	583	8	18	33
50	1.16	1.9	0.10	11	580	8	19	35
49	1.16	1.9	0.10	11	580	8	19	35
48	1.14	1.8	0.10	11	578	8	19	36
47	1.14	1.8	0.10	11	577	8	19	36
46	1.13	1.8	0.10	11	577	8	20	36
45	1.06	1.8	0.10	12	572	8	21	39
44	1.02	1.7	0.10	13	568	8	22	40
43	0.93	1.6	0.10	14	562	8	24	43
42	0.89	1.6	0.10	15	558	8	26	45
41	0.86	1.6	0.10	15	556	8	26	46
40	0.75	1.5	0.10	15	548	8	30	50
39	0.75	1.5	0.10	15	547	8	30	50
PISA Level IV = 545 (Strong Performers)								
38	0.67	1.4	0.10	19	541	8	32	53
37	0.64	1.3	0.10	19	539	8	33	54
36	0.56	1.3	0.10	21	533	8	35	57
35	0.51	1.2	0.09	22	528	8	37	60
34	0.45	1.2	0.09	23	524	8	39	62
33	0.39	1.1	0.10	25	519	8	41	64
32	0.33	1.0	0.10	26	515	8	43	66
31	0.30	1.0	0.10	27	513	8	44	67
30	0.22	0.9	0.10	29	506	8	47	70
29	0.19	0.9	0.10	30	504	8	47	71
28	0.11	0.8	0.10	33	498	8	50	73
27	0.08	0.8	0.09	33	495	8	51	74
26	0.02	0.7	0.09	35	491	8	53	76
25	-0.06	0.7	0.10	35	484	8	56	78
24	0.23	0.6	0.08	35	483	7	57	79
PISA Level III = 482 (Moderate Performers)								
23	-0.15	0.6	0.10	37	477	8	59	81
22	-0.18	0.5	0.10	42	475	8	60	82
21	-0.27	0.4	0.10	45	468	8	63	84

20	-0.32	0.4	0.10	47	464	8	65	85
19	-0.36	0.4	0.10	48	461	8	66	86
18	-0.45	0.3	0.10	52	454	8	69	88
17	-0.51	0.2	0.10	54	449	8	70	89
16	-0.61	0.1	0.10	58	441	8	73	91
15	-0.66	0.1	0.10	59	438	8	74	91
14	-0.70	0.0	0.10	61	434	8	76	92
13	-0.73	0.0	0.10	62	432	8	76	92
12	-0.81	-0.1	0.10	65	426	8	78	93
11	-0.86	-0.1	0.10	66	422	8	80	94
PISA Level II = 420 (Moderate Performers)								
10	-0.92	-0.2	0.11	68	418	8	81	95
9	-0.93	-0.2	0.10	69	416	8	81	95
8	-1.01	-0.3	0.11	71	410	8	83	96
7	-1.08	-0.4	0.11	73	405	9	84	96
6	-1.20	-0.5	0.11	77	395	9	87	97
5	-1.22	-0.5	0.11	77	394	9	87	97
4	-1.27	-0.6	0.11	79	390	9	88	97
3	-1.37	-0.7	0.12	81	382	9	90	98
2	-1.48	-0.8	0.12	84	374	9	91	98
1	-1.50	-0.8	0.12	84	372	9	91	99

Note: The 2006 PISA inverse cumulative percentages (percent reaching the PISA-equivalents) in mathematics were not available to the authors. Therefore, the distributions were estimated assuming a normal approximation (30 OECD countries, mean = 498, standard deviation = 92; Finland, mean = 548, standard deviation = 90). Any country or combination of countries that participated in 2006 PISA can be selected for comparisons. The result of the standard setting was to raise the Meets Proficiency cut-score (this is the performance standard used for federal accountability reporting) from page 19 in the ordered-item booklet to page 26.

Table 10: International Benchmarking With PISA for Hawaii, Reading, 2010

Ordered-Item Booklet Page	Item Parameter	Theta Associated With RP67	Standard Error of Theta Associated With RP67	Percent Reaching State Cut-Score	PISA-Equivalent Scaled Score	Standard Error of PISA-Equivalent	Percent Reaching OECD PISA-Equivalent	Percent Reaching Finland PISA-Equivalent
65	2.89	3.7	0.14	5	657	11	5	9
PISA Level V = 626 (Top Performers)								
64	2.13	2.8	0.11	11	592	8	16	29
63	1.96	2.7	0.10	13	578	8	19	35
62	1.76	2.5	0.10	15	562	8	24	42
61	1.76	2.5	0.10	15	562	8	24	42
60	1.71	2.4	0.10	16	558	8	25	45
59	1.70	2.4	0.10	16	558	8	26	45
58	1.69	2.4	0.10	16	557	8	26	45
57	1.69	2.4	0.10	17	557	8	26	45
56	1.66	2.4	0.10	17	554	8	27	47
PISA Level IV = 553 (Strong Performers)								
55	1.58	2.3	0.10	18	548	8	29	50
54	1.57	2.3	0.10	19	547	8	29	50
53	1.57	2.3	0.10	19	547	8	29	50
52	1.48	2.2	0.10	20	540	8	32	54
51	1.14	2.1	0.09	21	537	7	33	55
50	1.44	2.1	0.10	21	536	8	33	55
49	1.44	2.1	0.10	21	536	8	33	55
48	1.31	2.0	0.10	24	526	8	37	60
47	1.31	2.0	0.10	24	526	8	37	60
46	1.28	2.0	0.10	25	524	8	38	61
45	1.23	1.9	0.10	26	520	8	39	63
44	1.20	1.9	0.10	27	517	8	40	64
43	1.19	1.9	0.10	27	517	8	40	65
42	1.11	1.8	0.10	29	510	8	43	67
41	1.06	1.8	0.10	30	506	8	44	69
40	1.06	1.8	0.10	31	506	8	44	69
39	1.05	1.8	0.10	31	506	8	45	70
38	1.01	1.7	0.10	32	502	8	46	71
37	0.98	1.7	0.10	33	500	8	47	72
36	0.97	1.7	0.10	33	499	8	47	72
35	0.85	1.6	0.08	34	496	6	48	73
34	0.92	1.6	0.10	34	495	8	49	74
33	0.92	1.6	0.10	34	495	8	49	74
32	0.85	1.6	0.10	36	489	8	51	76
31	0.81	1.5	0.10	38	486	8	52	77
30	0.79	1.5	0.10	38	485	8	53	78
PISA Level III = 480 (Moderate Performers)								
29	0.71	1.4	0.10	41	479	8	55	80
28	0.69	1.4	0.10	41	477	8	56	81
27	0.64	1.4	0.10	43	473	8	57	82
26	0.61	1.3	0.10	44	471	8	58	83
25	0.61	1.3	0.10	44	470	8	59	83

24	0.60	1.3	0.10	44	470	8	59	83
23	0.57	1.3	0.10	45	467	8	60	84
22	0.51	1.2	0.10	47	462	8	62	85
21	0.48	1.2	0.10	48	460	8	63	86
20	0.42	1.1	0.10	50	455	8	64	87
19	1.09	1.1	0.08	50	454	6	65	87
18	0.32	1.0	0.10	53	447	8	67	89
17	0.32	1.0	0.10	54	447	8	67	89
16	-0.27	1.0	0.08	55	444	7	69	90
15	0.25	1.0	0.10	56	442	8	69	90
14	0.17	0.9	0.10	59	435	8	72	92
13	0.06	0.8	0.11	62	426	8	75	93
12	0.74	0.7	0.08	63	425	7	75	93
11	-0.06	0.7	0.11	66	417	9	77	95
10	-0.07	0.6	0.11	66	416	9	78	95
PISA Level II = 408 (Moderate Performers)								
9	-0.19	0.5	0.11	70	406	9	80	96
8	0.27	0.5	0.08	72	401	6	82	96
7	-0.26	0.4	0.11	72	400	9	82	96
6	-0.38	0.3	0.12	75	391	9	84	97
5	-0.46	0.3	0.12	77	385	9	86	98
4	-0.52	0.2	0.12	79	380	10	87	98
3	0.26	0.0	0.09	84	363	8	90	99
2	-0.79	-0.1	0.13	85	358	10	91	99
PISA Level I = 335 (Lowest Performers)								
1	-1.09	-0.4	0.14	90	334	11	94	100

Note: The 2006 PISA inverse cumulative percentages (percent reaching the PISA-equivalents) in reading were not available to the authors. Therefore, the distributions were estimated assuming a normal approximation (30 OECD countries, mean = 492, standard deviation = 100; Finland, mean = 547, standard deviation = 81). Any country or combination of countries that participated in 2006 PISA can be selected for comparisons. The result of the standard setting was to raise the Meets Proficiency cut-score (this is the performance standard used for federal accountability reporting) from page 6 in the ordered-item booklet to page 19.

Table 11: International Benchmarking With PISA for Delaware, Science, 2010

Ordered-Item Booklet Page	Item Parameter	Theta Associated With RP67	Standard Error of Theta Associated With RP67	Percent Reaching State Cut-Score	PISA-Equivalent Scaled Score	Standard Error of PISA-Equivalent	Percent Reaching OECD PISA-Equivalent	Percent Reaching Finland PISA-Equivalent
64	0.80	1.7	0.13	1	645	12	6	17
PISA Level V = 633 (Top Performers)								
63	0.58	1.3	0.10	3	604	10	14	32
62	0.54	1.3	0.10	4	601	10	14	33
61	0.50	1.2	0.10	4	597	10	15	35
60	0.48	1.2	0.10	4	595	10	16	35
59	0.48	1.2	0.10	4	595	10	16	35
58	0.42	1.1	0.10	5	589	10	17	38
57	0.38	1.1	0.10	5	586	10	18	39
56	0.38	1.1	0.10	5	585	9	19	40
55	0.19	0.9	0.10	7	568	9	24	48
54	0.16	0.9	0.10	8	565	9	25	49
53	0.15	0.9	0.10	8	564	9	25	50
52	0.14	0.9	0.10	8	563	9	25	50
51	0.13	0.8	0.10	8	562	9	26	50
50	-0.54	0.8	0.08	8	562	8	27	50
49	0.11	0.8	0.10	9	560	9	29	51
PISA Level IV = 559 (Strong Performers)								
48	0.08	0.8	0.10	9	558	9	29	52
47	0.05	0.8	0.10	10	555	9	29	54
46	0.03	0.7	0.10	10	553	9	29	55
45	0.00	0.7	0.10	10	550	9	30	56
44	-0.05	0.7	0.10	11	546	9	31	58
43	-0.09	0.6	0.10	12	542	9	33	60
42	-0.12	0.6	0.10	14	539	9	34	61
41	-0.12	0.6	0.09	14	539	9	34	61
40	-0.12	0.6	0.09	15	539	9	34	61
39	0.23	0.6	0.08	22	535	7	36	63
38	-0.18	0.5	0.09	22	533	9	36	64
37	-0.18	0.5	0.09	22	533	9	36	64
36	-0.19	0.5	0.10	22	533	9	36	64
35	-0.25	0.5	0.10	23	526	9	39	67
34	0.60	0.3	0.07	24	516	7	43	71
33	-0.38	0.3	0.09	24	514	9	44	72
32	-0.43	0.3	0.09	24	510	9	46	73
31	-0.44	0.3	0.09	24	509	9	46	73
30	-0.46	0.2	0.10	24	507	9	47	74
29	-0.53	0.2	0.09	26	501	9	50	76
28	-0.57	0.1	0.10	28	497	9	51	78
27	-0.60	0.1	0.09	29	494	9	53	79
26	-0.62	0.1	0.09	30	492	9	53	80
25	-0.62	0.1	0.10	30	492	9	53	80
24	-0.63	0.1	0.09	31	491	9	54	80

23	-0.64	0.1	0.09	31	491	9	54	80
22	-0.69	0.0	0.10	33	486	9	57	81
PISA Level III = 484 (Moderate Performers)								
21	-0.78	-0.1	0.10	39	477		60	84
20	-0.79	-0.1	0.10	40	477	9	60	84
19	-0.79	-0.1	0.10	54	476	9	60	84
18	-0.86	-0.1	0.10	54	470	9	62	86
17	-0.91	-0.2	0.10	55	465	9	64	87
16	-0.93	-0.2	0.10	55	463	9	65	88
15	-0.97	-0.3	0.10	56	460	9	66	88
14	-0.98	-0.3	0.10	56	459	9	67	89
13	-1.03	-0.3	0.10	56	454	9	69	90
12	-1.07	-0.4	0.10	57	450	9	70	91
11	-1.14	-0.4	0.10	57	444	9	72	92
10	-1.20	-0.5	0.10	86	439	9	74	93
9	-1.21	-0.5	0.10	86	437	10	75	93
8	-1.29	-0.6	0.10	86	430	10	77	94
7	-1.32	-0.6	0.10	86	427	10	78	94
6	-0.91	-0.6	0.08	86	424	7	79	95
5	-1.01	-0.7	0.08	86	419	8	80	95
4	-1.49	-0.8	0.11	87	411	10	83	96
3	-1.50	-0.8	0.11	87	410	10	83	96
PISA Level II = 409 (Moderate Performers)								
2	-1.65	-0.9	0.11	87	396	11	86	97
1	-1.79	-1.1	0.11	88	383	11	89	98

Note: The 2006 PISA inverse cumulative percentages (percent reaching the PISA-equivalents) in science were not available to the authors. Therefore, the distributions were estimated assuming a normal approximation (30 OECD countries, mean = 500, standard deviation = 95; Finland, mean = 563, standard deviation = 86). Any country or combination of countries that participated in 2006 PISA can be selected for comparisons. The result of the standard setting was to set the Meets Standard cut-score (this is the performance standard used for federal accountability reporting) on page 21 in the ordered-item booklet. There was no prior standard on the test.

Table 12: International Benchmarking With PISA for Delaware, Mathematics, 2010

Ordered-Item Booklet Page	Item Parameter	Theta Associated With RP67	Standard Error of Theta Associated With RP67	Percent Reaching State Cut-Score	PISA-Equivalent Scaled Score	Standard Error of PISA-Equivalent	Percent Reaching OECD PISA-Equivalent	Percent Reaching Finland PISA-Equivalent
64	3.75	4.5	0.15	2	669	11	3	7
63	3.49	4.2	0.13	3	649	10	5	11
62	3.37	4.1	0.13	4	640	10	6	13
61	3.29	4.0	0.13	4	634	10	7	14
60	3.17	3.9	0.12	4	624	10	9	17
PISA Level V = 607 (Top Performers)								
59	2.94	3.6	0.09	6	606	7	12	24
58	2.87	3.6	0.12	6	601	9	13	26
57	2.80	3.5	0.12	7	596	9	14	28
56	2.74	3.4	0.11	7	591	9	16	30
55	2.71	3.4	0.08	7	588	7	16	31
54	2.65	3.4	0.08	8	584	6	17	33
53	2.61	3.3	0.11	8	581	9	18	34
52	2.55	3.3	0.11	9	576	9	20	36
51	2.48	3.2	0.11	10	570	9	22	39
50	2.42	3.1	0.11	10	566	8	23	41
49	2.34	3.1	0.11	11	560	9	25	44
48	2.29	3.0	0.11	12	555	8	27	47
47	2.26	3.0	0.10	12	553	8	27	48
46	2.20	2.9	0.11	12	549	8	29	50
45	2.18	2.9	0.10	12	547	8	30	50
44	2.17	2.9	0.10	12	546	8	30	51
PISA Level IV = 545 (Strong Performers)								
43	2.11	2.8	0.10	15	541	8	32	53
42	2.07	2.8	0.10	15	538	8	33	55
41	2.04	2.7	0.10	16	536	8	34	56
40	1.99	2.7	0.10	17	533	8	35	57
39	1.97	2.7	0.10	28	531	8	36	58
38	1.94	2.7	0.07	28	529	6	37	59
37	1.94	2.6	0.10	28	528	8	37	60
36	1.89	2.6	0.10	28	525	8	38	61
35	1.89	2.6	0.10	29	524	8	39	62
34	1.85	2.6	0.10	29	522	8	40	63
33	1.82	2.5	0.10	29	519	8	41	64
32	1.77	2.5	0.10	29	515	8	43	66
31	1.74	2.4	0.10	29	513	8	44	67
30	1.68	2.4	0.10	29	508	8	46	69
29	1.65	2.4	0.10	29	506	8	47	70
28	1.63	2.3	0.10	30	504	8	47	71
27	1.59	2.3	0.10	30	501	8	49	72
26	1.55	2.3	0.10	30	498	8	50	73
25	1.51	2.2	0.10	30	495	8	51	74
24	1.49	2.2	0.10	30	493	8	52	75
23	1.45	2.2	0.10	30	490	8	53	76

22	1.41	2.1	0.10	30	487	8	55	77
21	1.39	2.1	0.10	30	485	8	56	78
PISA Level III = 482 (Moderate Performers)								
20	1.32	2.0	0.10	34	480	8	58	80
19	1.25	2.0	0.10	37	474	8	60	82
18	1.20	1.9	0.10	38	471	8	62	83
17	1.16	1.9	0.10	40	468	8	63	84
16	1.11	1.8	0.10	42	463	8	65	85
15	1.08	1.8	0.10	43	461	8	66	86
14	1.07	1.8	0.07	43	461	6	66	86
13	1.03	1.7	0.10	44	458	8	67	87
12	1.02	1.7	0.10	45	456	8	68	87
11	0.96	1.7	0.10	48	452	8	69	88
10	0.87	1.6	0.10	66	445	8	72	90
9	0.84	1.6	0.07	67	443	6	73	90
8	0.75	1.5	0.11	67	435	9	75	92
7	0.75	1.5	0.11	67	435	8	75	92
6	0.70	1.4	0.11	67	432	9	76	92
5	0.63	1.3	0.11	68	426	9	78	93
4	0.57	1.3	0.11	68	421	9	80	94
PISA Level II = 420 (Moderate Performers)								
3	0.37	1.1	0.11	72	406	9	84	96
2	0.26	1.0	0.11	95	398	9	86	97
1	0.05	0.8	0.12	95	381	9	90	98

Note: The 2006 PISA inverse cumulative percentages (percent reaching the PISA-equivalents) in mathematics were not available to the authors. Therefore, the distributions were estimated assuming a normal approximation (30 OECD countries, mean = 498, standard deviation = 92; Finland, mean = 548, standard deviation = 90). Any country or combination of countries that participated in 2006 PISA can be selected for comparisons. The result of the standard setting was to raise the Meets Standard cut-score (this is the performance standard used for federal accountability reporting) from page 10 in the ordered-item booklet to page 12.

Table 13: International Benchmarking With PISA for Delaware, Reading, 2010

Ordered-Item Booklet Page	Item Parameter	Theta Associated With RP67	Standard Error of Theta Associated With RP67	Percent Reaching State Cut-Score	PISA-Equivalent Scaled Score	Standard Error of PISA-Equivalent	Percent Reaching OECD PISA-Equivalent	Percent Reaching Finland PISA-Equivalent
63	2.51	3.2	0.10	10	587	8	17	31
62	2.33	3.0	0.10	12	574	8	21	37
61	2.24	2.9	0.10	13	566	8	23	41
60	2.18	2.9	0.10	14	561	8	25	43
59	2.14	2.8	0.10	16	558	8	29	45
PISA Level IV = 553 (Strong Performers)								
58	1.93	2.8	0.09	16	551	8	29	48
57	2.01	2.7	0.10	17	547	8	29	50
56	1.75	2.5	0.10	22	527	8	36	60
55	1.75	2.5	0.10	22	527	8	36	60
54	1.69	2.4	0.10	24	522	8	38	62
53	1.62	2.3	0.10	26	516	8	40	65
52	0.94	2.2	0.09	28	509	7	43	68
51	1.51	2.2	0.10	28	507	8	44	69
50	1.49	2.2	0.10	29	506	8	44	69
49	1.45	2.2	0.10	30	503	8	46	71
48	1.41	2.1	0.10	31	500	8	47	72
47	1.40	2.1	0.10	32	498	8	47	73
46	1.37	2.1	0.10	32	497	8	48	73
45	1.36	2.1	0.10	33	495	8	49	74
44	1.34	2.1	0.10	33	494	8	49	74
43	1.32	2.0	0.10	34	492	8	50	75
42	1.30	2.0	0.10	35	491	8	51	76
41	1.28	2.0	0.10	35	489	8	51	76
40	1.25	2.0	0.10	36	487	8	52	77
39	1.24	1.9	0.10	36	486	8	52	77
38	1.17	1.9	0.10	38	481	8	57	79
PISA Level III = 480 (Moderate Performers)								
37	1.00	1.7	0.10	44	467	8	60	84
36	1.00	1.7	0.10	44	467	8	60	84
35	0.96	1.7	0.10	45	464	8	61	85
34	0.96	1.7	0.10	45	464	8	61	85
33	0.96	1.7	0.10	45	463	8	61	85
32	0.94	1.6	0.10	46	462	8	62	85
31	2.03	1.6	0.08	46	461	7	62	86
30	0.88	1.6	0.10	48	458	8	63	87
29	0.88	1.6	0.07	48	457	6	64	87
28	-0.17	1.5	0.08	50	452	7	65	88
27	0.74	1.4	0.10	53	446	8	68	89
26	0.74	1.4	0.10	53	446	8	68	89
25	0.72	1.4	0.07	53	445	6	68	90
24	0.72	1.4	0.10	53	444	8	68	90
23	0.71	1.4	0.10	54	443	8	69	90
22	0.69	1.4	0.10	54	442	8	69	90

21	0.63	1.3	0.10	56	437	8	71	91
20	0.60	1.3	0.10	57	435	8	72	92
19	0.58	1.3	0.10	58	433	8	72	92
18	0.57	1.3	0.10	58	432	8	73	92
17	0.54	1.3	0.10	59	430	8	73	93
16	0.54	1.2	0.10	59	430	9	73	93
15	2.05	1.2	0.08	62	424	7	75	93
14	0.45	1.2	0.10	62	423	8	76	94
13	0.43	1.1	0.11	63	421	9	76	94
12	0.40	1.1	0.07	64	418	6	77	94
11	0.37	1.1	0.11	65	417	9	77	95
10	0.71	1.1	0.09	65	416	7	78	95
9	0.33	1.0	0.11	66	413	9	79	95
8	0.30	1.0	0.11	67	411	9	79	95
PISA Level II = 408 (Moderate Performers)								
7	0.21	0.9	0.11	70	404	9	81	96
6	0.19	0.9	0.07	71	402	6	82	96
5	0.15	0.9	0.11	72	398	9	83	97
4	-0.02	0.7	0.11	77	385	9	86	98
3	-0.11	0.6	0.11	79	378	9	87	98
2	-0.37	0.3	0.12	86	357	10	91	99
1	-0.49	0.2	0.12	88	347	10	93	99

Note: The 2006 PISA inverse cumulative percentages (percent reaching the PISA-equivalents) in reading were not available to the authors. Therefore, the distributions were estimated assuming a normal approximation (30 OECD countries, mean = 492, standard deviation = 100; Finland, mean = 547, standard deviation = 81). Any country or combination of countries that participated in 2006 PISA can be selected for comparisons. The result of the standard setting was to raise the Meets Standard cut-score (this is the performance standard used for federal accountability reporting) from page 12 in the ordered-item booklet to page 31.

Table 14: International Benchmarking With PISA for Oregon, Mathematics, 2010

Ordered-Item Booklet Page	Item Parameter	Theta Associated With RP67	Standard Error of Theta Associated With RP67	Percent Reaching State Cut-Score	PISA-Equivalent Scaled Score	Standard Error of PISA-Equivalent	Percent Reaching OECD PISA-Equivalent	Percent Reaching Finland PISA-Equivalent
67	6.45	7.2	0.18	0	772	14	0	0
66	6.00	6.7	0.13	0	737	10	0	1
65	5.90	6.6	0.15	0	730	12	1	1
64	5.80	6.5	0.09	1	722	7	1	2
63	5.70	6.4	0.14	1	714	11	1	2
62	5.60	6.3	0.09	1	706	7	1	3
61	5.50	6.2	0.08	1	698	6	1	3
60	5.40	6.1	0.13	1	691	10	2	4
59	4.46	6.0	0.16	1	686	12	2	4
58	5.30	6.0	0.08	1	683	7	2	5
57	5.20	5.9	0.12	2	675	9	3	6
PISA Level VI = 669 (Top Performers)								
56	5.10	5.8	0.05	2	667	4	3	7
55	5.00	5.7	0.11	2	660	9	4	8
54	6.38	5.6	0.13	3	654	10	4	10
53	4.90	5.6	0.11	3	652	8	5	10
52	4.80	5.5	0.11	3	644	8	6	12
51	4.70	5.4	0.05	4	636	4	7	14
50	4.60	5.3	0.05	4	628	4	8	16
49	4.50	5.2	0.06	5	621	5	9	18
48	4.40	5.1	0.10	6	613	8	11	21
47	3.96	5.1	0.09	6	609	7	11	23
PISA Level V = 607 (Top Performers)								
46	4.30	5.0	0.10	7	605	7	12	27
45	4.20	4.9	0.09	8	597	7	14	31
44	4.10	4.8	0.09	10	589	7	16	34
43	4.00	4.7	0.03	11	582	3	18	37
42	3.90	4.6	0.03	13	574	2	20	37
41	3.90	4.6	0.09	13	574	7	20	41
40	3.80	4.5	0.03	16	566	2	23	41
39	3.80	4.5	0.09	16	566	7	23	45
38	3.70	4.4	0.09	18	558	7	26	49
37	3.60	4.3	0.09	21	550	7	29	49
36	3.60	4.3	0.08	21	550	6	29	51
35	4.23	4.2	0.07	24	545	5	30	52
PISA Level IV = 545 (Strong Performers)								
34	3.50	4.2	0.09	24	543	7	31	56
33	3.40	4.1	0.03	28	535	2	34	60
32	3.30	4.0	0.09	32	527	7	38	64
31	3.20	3.9	0.09	37	519	7	41	64
30	3.20	3.9	0.09	37	519	7	41	67
29	3.10	3.8	0.03	43	512	3	44	67
28	3.10	3.8	0.03	43	512	2	44	71
27	3.00	3.7	0.09	49	504	7	47	71

26	3.00	3.7	0.09	49	504	7	47	74
25	2.90	3.6	0.03	56	496	2	51	74
24	2.90	3.6	0.03	56	496	2	51	77
23	2.80	3.5	0.09	60	488	7	54	77
22	2.80	3.5	0.09	60	488	7	54	80
PISA Level III = 482 (Moderate Performers)								
21	2.70	3.4	0.03	63	480	2	58	80
20	2.70	3.4	0.09	63	480	7	58	82
19	2.60	3.3	0.03	67	473	2	61	85
18	2.50	3.2	0.03	70	465	3	64	85
17	2.50	3.2	0.09	70	465	7	64	87
16	2.40	3.1	0.04	74	457	3	67	87
15	2.40	3.1	0.09	74	457	7	67	89
14	2.30	3.0	0.03	77	449	3	70	89
13	2.30	3.0	0.09	77	449	7	70	91
12	2.20	2.9	0.04	80	441	3	73	92
11	2.40	2.8	0.07	83	435	6	75	92
10	2.10	2.8	0.10	83	434	7	76	93
9	2.00	2.7	0.06	86	426	5	78	95
PISA Level II = 420 (Moderate Performers)								
8	1.90	2.6	0.10	88	418	8	81	96
7	1.80	2.5	0.10	90	410	8	83	96
6	1.70	2.4	0.11	92	402	8	85	97
5	1.60	2.3	0.11	94	395	8	87	98
4	1.50	2.2	0.11	95	387	9	89	98
3	1.40	2.1	0.11	96	379	9	90	99
2	1.30	2.0	0.12	97	371	9	92	99
PISA Level I = 358 (Lowest Performers)								
1	0.90	1.6	0.05	99	340	4	96	100

Note: The 2006 PISA inverse cumulative percentages (percent reaching the PISA-equivalents) in mathematics were not available to the authors. Therefore, the distributions were estimated assuming a normal approximation (30 OECD countries, mean = 498, standard deviation = 92; Finland, mean = 548, standard deviation = 90). Any country or combination of countries that participated in 2006 PISA can be selected for comparisons. The result of the standard setting was to raise the Meets cut-score (this is the performance standard used for federal accountability reporting) from between pages 25 and 26 in the ordered-item booklet to between pages 27 and 28.

Table 15: International Benchmarking With PISA for Oregon, Reading, 2010

Ordered-Item Booklet Page	Item Parameter	Theta Associated With RP67	Standard Error of Theta Associated With RP67	Percent Reaching State Cut-Score	PISA Equivalent	Standard Error PISA Equivalent	Percent Reaching OECD PISA-Equivalent	Percent Reaching Finland PISA-Equivalent
65	5.1	5.8	0.10	2	667	8	4	7
64	5.0	5.7	0.06	2	661	5	5	8
63	4.9	5.6	0.03	3	648	3	6	11
62	4.8	5.5	0.06	3	646	5	6	11
61	4.6	5.3	0.03	5	627	3	9	16
PISA Level V = 626 (Top Performers)								
60	4.5	5.2	0.08	7	619	7	10	19
59	4.5	5.2	0.03	7	619	3	10	19
58	4.4	5.1	0.05	8	611	4	12	21
57	4.4	5.1	0.10	8	611	8	12	21
56	4.3	5.0	0.03	10	603	3	13	24
55	4.2	4.9	0.02	13	596	2	15	27
54	4.1	4.8	0.06	15	586	5	17	31
53	4.1	4.8	0.07	15	586	5	17	31
52	4.0	4.7	0.03	18	580	3	19	34
51	4.0	4.7	0.05	18	580	4	19	34
50	3.9	4.6	0.03	21	571	3	21	38
49	3.9	4.6	0.03	21	571	2	21	38
48	3.8	4.5	0.05	24	563	4	24	42
47	3.7	4.4	0.03	28	555	3	26	46
46	3.7	4.4	0.03	28	555	2	26	46
PISA Level IV = 553 (Strong Performers)								
45	3.6	4.3	0.05	32	547	4	29	50
44	3.6	4.3	0.03	32	547	3	29	50
43	3.5	4.2	0.05	37	539	4	32	54
42	3.5	4.2	0.04	37	539	3	32	54
41	3.4	4.1	0.04	42	531	4	35	58
40	3.3	4.0	0.05	48	523	4	38	61
39	3.3	4.0	0.03	48	523	3	38	61
38	3.2	3.9	0.04	53	515	3	41	65
37	3.2	3.9	0.04	53	515	3	41	65
36	3.1	3.8	0.04	59	507	4	44	69
35	3.1	3.8	0.02	59	507	2	44	69
34	3.0	3.7	0.04	66	499	3	47	72
33	3.0	3.7	0.03	66	499	3	47	72
32	2.9	3.6	0.04	72	491	3	50	76
31	2.9	3.6	0.09	72	491	7	50	76
30	2.8	3.5	0.05	75	483	4	53	78
PISA Level III = 480 (Moderate Performers)								
29	2.7	3.4	0.08	78	475	6	57	81
28	2.7	3.4	0.08	78	475	6	57	81
27	2.6	3.3	0.10	80	467	8	60	84
26	2.5	3.2	0.04	83	459	3	63	86
25	2.5	3.2	0.05	83	459	4	63	86

24	2.4	3.1	0.04	85	451	3	66	88
23	2.3	3.0	0.03	87	443	3	69	90
22	2.3	3.0	0.04	87	443	3	69	90
21	2.2	2.9	0.06	89	435	5	71	92
20	2.1	2.8	0.05	90	427	4	74	93
19	2.1	2.8	0.05	90	427	4	74	93
18	2.0	2.7	0.07	92	419	5	77	94
17	1.9	2.6	0.13	93	410	11	79	95
16	1.9	2.6	0.07	93	410	5	79	95
PISA Level II = 408 (Moderate Performers)								
15	1.8	2.5	0.05	94	403	4	81	96
14	1.7	2.4	0.05	95	392	4	84	97
13	1.7	2.4	0.04	95	392	4	84	97
12	1.6	2.3	0.05	96	387	4	85	98
11	1.6	2.3	0.07	96	387	6	85	98
10	1.5	2.2	0.06	97	379	5	87	98
9	1.4	2.1	0.05	97	368	4	89	99
8	1.4	2.1	0.07	97	368	6	89	99
7	1.3	2.0	0.09	98	368	7	89	99
6	1.2	1.9	0.05	98	352	4	92	99
5	1.1	1.8	0.06	99	344	5	93	99
4	1.1	1.8	0.17	99	344	13	93	99
3	1.0	1.7	0.07	99	339	6	94	99
2	0.9	1.6	0.11	99	336	9	94	100
PISA Level I = 335 (Lowest Performers)								
1	0.8	1.5	0.10	99	323	8	95	100

Note: The 2006 PISA inverse cumulative percentages (percent reaching the PISA-equivalents) in reading were not available to the authors. Therefore, the distributions were estimated assuming a normal approximation (30 OECD countries, mean = 492, standard deviation = 100; Finland, mean = 547, standard deviation = 81). Any country or combination of countries that participated in 2006 PISA can be selected for comparisons. The result of the standard setting did not change the prior "Meets" cut-score (this is the performance standard used for federal accountability reporting). The performance standard remained between pages 32 - 33 in the ordered-item booklet.

Appendix:

Derivation of the Standard Error of the Theta Associated With the Response Probability (RP) for the Rasch Model

This appendix outlines the approach to estimate the theta score for a given response probability (*RP*) p, given a set of item parameter estimates and their standard errors and co-variances.

1. Dichotomous Item

In case of dichotomous or multiple choice item, the response probability of correct response given the item parameters b is

$$\Pr_1(\theta) = \Pr(x = 1 | b) = \frac{E\exp(\theta - b)}{1 + E\exp(\theta - b)}.$$

For given RP value,

$$\Pr_1(\theta) = \Pr(x = 1 | b) = \frac{E\exp(\theta - b)}{1 + E\exp(\theta - b)} = p,$$

hence,

$$\theta = b + \log\frac{p}{1 - p}.$$

The variance of θ can be estimated using delta method,

$$Var(\theta) = \left(\frac{\partial\theta}{\partial b}\right)^2 Var(b) = \sigma_b^2.$$

2. Open-Ended Item Using Partial Credit Model

Suppose that the item has K score categories (e.g., $0,1,2,\ldots,K-1$), and the item parameters are $b_1,\ldots,b_{K-1}$ using Mastersô(1982) parameterization, the probability of response in category j can be computed as

$$\Pr_j(\theta) = \Pr(x = j \mid b_1,\ldots,b_{K-1},\theta) = \frac{\exp\left[\sum_{i=1}^{j}(\theta - b_i)\right]}{1+\sum_{k=1}^{K-1} Exp(\theta - b_i)} \qquad \text{for } j = 0,1,2,\ldots,K-1.$$

For given p and response category $j \geq 1$, the RP score θ_j based on RP p satisfies

$$\sum_{m=j}^{K-1}\Pr_m(\theta_j) = \frac{\sum_{m=j}^{K-1}\exp\left[\sum_{i=1}^{m}(\theta - b_i)\right]}{1+\sum_{k=1}^{K-1}\exp\left[\sum_{i=1}^{k}(\theta - b_i)\right]} = p$$

and is obtained by solving the above equation numerically.

Define: $t_j = \exp(\theta_j), h_k = \exp\left[-\sum_{i=1}^{k} b_i\right]$, then θ_j or t_j satisfies

$$1+\sum_{m=1}^{j-1} h_m t_j^m - \frac{1-p}{p}\sum_{m=j}^{K-1} h_m t_j^m = 0,\ t_j \geq 0$$

The variance of θ_j can be estimated by using delta method:

$$Var\left(\theta_j\right)=\left(\frac{\partial\theta_j}{\partial b_1}\quad\cdots\quad\frac{\partial\theta_j}{\partial b_{K-1}}\right)Var\begin{pmatrix}b_1\\\vdots\\b_{K-1}\end{pmatrix}\left(\frac{\partial\theta_j}{\partial b_1}\quad\cdots\quad\frac{\partial\theta_j}{\partial b_{K-1}}\right)'$$
$$=\sum_{k=1}^{K-1}\left(\frac{\partial\theta_j}{\partial b_k}\right)^2\sigma_{b_k}^2+2\sum_{k_1=1}^{K-1}\sum_{k_2=k_1+1}^{K-1}\left(\frac{\partial\theta_j}{\partial b_{k_1}}\frac{\partial\theta_j}{\partial b_{k_2}}\right)\sigma_{b_{k_1},b_{k_2}}$$

where

$$\frac{\partial\theta_j}{\partial b_k}=\begin{cases}\dfrac{\sum_{m=k}^{j-1}h_m t_j^m-\frac{1-p}{p}\sum_{m=j}^{K-1}h_m t_j^m}{\sum_{m=1}^{j-1}h_m m t_j^m-\frac{1-p}{p}\sum_{m=j}^{K-1}h_m m t_j^m} & \text{if } k\le j-1\\[2ex]\dfrac{-\frac{1-p}{p}\sum_{m=k}^{K-1}h_m t_j^m}{\sum_{m=1}^{j-1}h_m m t_j^m-\frac{1-p}{p}\sum_{m=j}^{K-1}h_m m t_j^m} & \text{if } k>j-1\end{cases}$$

References

Adams, R. (Ed.). (2009). *PISA 2006 Technical Report.* Paris: Organization for Economic Cooperation and Development (OECD).

Koretz, D. M., Bertenthal, M. W., & Green, B. F. (Eds.). (1999). *Embedding questions: The pursuit of a common measure in uncommon tests*. Washington, DC: National Academies Press.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika,* 47, 149ï 174.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives*. Mahwah, NJ: Erlbaum.

Thissen, D. (2007). Linking assessments based on aggregate reporting: Background and issues. In J. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 287ï 312). New York: Springer.