

Technical Analysis



PENNSYLVANIA SYSTEM OF SCHOOL ASSESSMENT

2004 WRITING PSSA GRADE ELEVEN

DATA RECOGNITION CORPORATION

**RONALD MEAD, PH.D.
SENIOR PSYCHOMETRICIAN**

**RICHARD M. SMITH, PH.D.
SENIOR PSYCHOMETRICIAN**

**ANDREW SWANLUND
STATISTICAL ANALYST**

**KRISTINE VAN BEUSEKOM
STATISTICAL ANALYST**

DECEMBER 2004

THE PENNSYLVANIA SYSTEM OF SCHOOL ASSESSMENT 2004

TABLE OF CONTENTS

Description	1
Recent Changes	1
Scoring	1
Table 1: Inter-Rater Agreement for Writing Tasks by Domain	2
Table 2: Reader and Domain Correlations	3
Table 3: Grade 11 Writing Summary Statistics	4
Table 4: Domain Level Means and Correlations	4
Table 5: Scoring Category Level Statistics	5
Calibration and Scaling	5
Equating	6
Table 6: Winsteps Calibration for 2004 Grade 11 Writing	7
Table 7: Step Logit Difficulties	8
Scaled Scores	8
Table 8: Raw Score to Logit Conversion Table	10
Table 9: Raw Score to Scaled Score Conversion Table	11
Student Level Results	12
Table 10: Summary Statistics for the Norming Group	12
Table 11: PSSA 2004 Grade 11 Writing Performance Levels	12
Indicators of Consistency	12
Table 12: Livingston Coefficient of Consistency at Each Performance Standard	13
School Level Results	13
Table 13: School Level Scores	13
Definition of Grade Eleven Writing Variable	14
Table 14: Grade 11 Writing	15
Appendix A: Winsteps Output Tables	A-1
Table 1: Grade 11 Item Calibrations	A-1
Raw Score/Logit Conversion Tables	
Table 2: Winsteps Category to Raw Score Conversion Domain Level	A-2
Table 3: Winsteps Category to Raw Score Conversion Prompt Level	A-2
Table 4: Winsteps Category to Raw Score Conversion Total Test	A-3
Table 5: Grade 11 Raw Score/Logit Conversion Table Total Test	A-4
Table 6: Grade 11 Raw Score/Logit Conversion Table Prompt 1 (Narrative)	A-5

Table 7: Grade11 Raw Score/Logit Conversion Table Prompt 2 (Persuasive)	A-5
Table 8: Grade11 Raw Score/Logit Conversion Table Prompt 3 (Informational)	A-5
Table 9: Grade11 Raw Score/Logit Conversion Table Domain 1 (Focus)	A-6
Table 10: Grade11 Raw Score/Logit Conversion Table Domain 2 (Content)	A-6
Table 11: Grade11 Raw Score/Logit Conversion Table Domain 3 (Organization)	A-6
Table 12: Grade11 Raw Score/Logit Conversion Table Domain 4 (Style)	A-6
Table 13: Grade11 Raw Score/Logit Conversion Table Domain 5 (Conventions)	A-7

Threshold Difficulties

Table 14: Prompt 1 Domain 1	A-7
Table 15: Prompt 1 Domain 2	A-8
Table 16: Prompt 1 Domain 3	A-8
Table 17: Prompt 1 Domain 4	A-9
Table 18: Prompt 1 Domain 5	A-9
Table 19: Prompt 2 Domain 1	A-10
Table 20: Prompt 2 Domain 2	A-10
Table 21: Prompt 2 Domain 3	A-11
Table 22: Prompt 2 Domain 4	A-11
Table 23: Prompt 2 Domain 5	A-12
Table 24: Prompt 3 Domain 1	A-12
Table 25: Prompt 3 Domain 2	A-13
Table 26: Prompt 3 Domain 3	A-13
Table 27: Prompt 3 Domain 4	A-14
Table 28: Prompt 3 Domain 5	A-14

DESCRIPTION

The PSSA writing assessment was administered in February 2004 to approximately 131,000 Grade 11 students. All students were expected to participate; private schools and home school students were permitted to participate on a volunteer basis.

The assessment consisted of three writing prompts, one prompt at each mode: Narrative, Persuasive and Informational. The Narrative prompt asked students to describe possible incidents of a first day on the job and to give reasons why they would stay or leave the job after their first day. The Persuasive prompt asked students to persuade the school board to accept their ideas on whether or not extracurricular activities are important to a student's total educational program. The final prompt was Informational and asked students to explain a concern that affects their neighborhood and to state why this issue is important to the neighborhood.

The hand-written responses were packeted and scored using standard DRC training and scoring procedures. Responses were all read twice with resolution. All students scoring within 5 points below the proficiency cutpoint were rescored.

Scores were reported in a scaled score metric that is analogous to those used for reading and mathematics in grades five, eight and eleven. As with the other grades, the scaled score metric was designed so that the 2001 base year mean school-level scaled score was 1300 and the standard deviation of the school-level scores was 100.

RECENT CHANGES

A minimum scaled score of 700 was implemented for all PSSA Reading, Math and Writing exams beginning in 2002. This minimum is applied for all PSSA scales. This change affects very few students; however, several schools felt they were being penalized by the presence of some special needs students who took the regular assessment but scored very low. The change was made to reduce the impact of these students on the overall school score. For Grade 11 Writing, there were fewer than 0.7% of the students statewide who received scores of 700.

Performance Levels of Below Basic, Basic, Proficient and Advanced were established by a standards setting committee of Pennsylvania teachers. Performance levels were reported for the first time beginning with the 2002 Grade 11 Writing assessment. Data is reported as the percent in each school and statewide that fall into each of the four levels. These percentages are an important indicator of the status and progress of schools.

Equating from the 2004 to the 2003 assessment was done by means of a common prompt. It is essential for equating across years that there be a common element in both assessments. For direct writing assessments, there is no alternative to re-using a prompt. At the current stage of development for the grade 11 assessment, this means using a prompt in the current year that was used in the previous year. The Persuasive prompt from 2003 was re-used on the 2004 assessment to provide a means of equating across years.

SCORING

Each writing sample was scored by two readers using a domain scoring model with five domains and a four-point rubric.

2003 PSSA Grade Eleven Writing

For a *scorable* paper, the final scale for a paper ranged from ten (all ones for the five domains from the two readers) to forty (all fours for the five domains from the two readers). A paper was deemed *nonscorable* if it was blank with no explanation, illegible, or deliberately off-topic. A nonscorable paper was given scores of zero for all domains.

A paper was considered a *non-attempt* if it was blank with a suitable explanation. The explanations included *due to absence* and *by parental request*. These papers were *not* given scores.

Responses were all read twice with resolution. For non-adjacent domain scores, final scores were calculated as follows:

- If the 3rd reader's score matches either of the original scores, then the 3rd reader's score plus the matching score (or the 3rd reader's score times two) is final.
- If the 3rd reader's score is adjacent to the scores of both reader 1 and reader 2, then the sum of the 3rd reader's score with that of the higher of the original two non-adjacent scores is final.
- If the 3rd reader's score is adjacent to one of the original two scores (but not both), then the final score is the sum of the 3rd reader's score and the original score it is adjacent to.

The total agreement indicates that nearly 100% of the readers rated papers similarly. Table 1 shows the inter-reader agreement for writing tasks at each domain.

Table 1: Inter-Reader Agreement for Writing Tasks by Domain Based on All Students

Prompt 1	Focus	Content	Organization	Style	Conventions
Exact	75.0%	73.0%	74.0%	73.0%	72.0%
Adjacent	25.0%	27.0%	26.0%	27.0%	28.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%
Prompt 2	Focus	Content	Organization	Style	Conventions
Exact	70.0%	71.0%	72.0%	69.0%	67.0%
Adjacent	29.0%	29.0%	28.0%	30.0%	32.0%
Total	99.0%	100.0%	100.0%	99.0%	99.0%
Prompt 3	Focus	Content	Organization	Style	Conventions
Exact	68.0%	72.0%	71.0%	70.0%	66.0%
Adjacent	32.0%	28.0%	28.0%	29.0%	33.0%
Total	99.0%	100.0%	99.0%	99.0%	99.0%

All students scoring within 5 raw score points below the proficiency cutpoint were rescored. Each of the three essays was read a third time and the two highest scores were reported. Ten percent of these essays were read again for quality assurance purposes.

Table 2 shows the correlations among the domains and between the readers. In general, the correlations among domains for a single reader are on the order of 0.7 to 0.8, while correlations among domains across two readers are between 0.6 and 0.7. The type of writing does not seem to be important in the correlations.

2003 PSSA Grade Eleven Writing

The correlation between two readers of the same paper is approximately 0.80 for each mode. While not shown in the table, the correlations at the prompt level between two papers are similar to the correlations between readers.

Table 2: Reader and Domain Correlations Based on All Students

Prompt 1: Narrative												
	R1D1	R1D2	R1D3	R1D4	R1D5	R1 Total	R2D1	R2D2	R2D3	R2D4	R2D5	R2 Total
R1D1	1	0.829	0.789	0.758	0.739	0.901	0.674	0.667	0.653	0.642	0.627	0.715
R1R2		1.000	0.827	0.783	0.744	0.919	0.670	0.716	0.683	0.665	0.644	0.741
R1D3			1.000	0.798	0.763	0.917	0.658	0.683	0.694	0.665	0.648	0.735
R1D4				1.000	0.853	0.921	0.646	0.666	0.664	0.690	0.672	0.733
R1D5					1.000	0.900	0.634	0.647	0.649	0.674	0.686	0.722
R1 Total						1.000	0.720	0.742	0.734	0.732	0.719	0.800
R2D1							1.000	0.831	0.792	0.764	0.742	0.903
R2D2								1.000	0.823	0.782	0.743	0.918
R2D3									1.000	0.797	0.761	0.915
R2D4										1.000	0.856	0.922
R2D5											1.000	0.901
R2 Total												1.000
Prompt 2: Persuasive												
	R1D1	R1D2	R1D3	R1D4	R1D5	R1 Total	R2D1	R2D2	R2D3	R2D4	R2D5	R2 Total
R1D1	1.000	0.825	0.808	0.761	0.707	0.904	0.660	0.661	0.650	0.637	0.598	0.708
R1R2		1.000	0.847	0.777	0.702	0.917	0.663	0.706	0.676	0.658	0.609	0.732
R1D3			1.000	0.800	0.735	0.924	0.649	0.674	0.681	0.657	0.616	0.724
R1D4				1.000	0.784	0.912	0.636	0.656	0.657	0.686	0.639	0.723
R1D5					1.000	0.869	0.597	0.608	0.618	0.640	0.666	0.691
R1 Total						1.000	0.708	0.730	0.725	0.725	0.692	0.791
R2D1							1.000	0.828	0.811	0.764	0.706	0.905
R2D2								1.000	0.849	0.776	0.701	0.918
R2D3									1.000	0.799	0.735	0.925
R2D4										1.000	0.786	0.912
R2D5											1.000	0.869
R2 Total												1.000
Prompt 3: Informational												
	R1D1	R1D2	R1D3	R1D4	R1D5	R1 Total	R2D1	R2D2	R2D3	R2D4	R2D5	R2 Total
R1D1	1.000	0.753	0.764	0.711	0.678	0.875	0.654	0.633	0.626	0.609	0.585	0.697
R1R2		1.000	0.794	0.751	0.691	0.895	0.639	0.721	0.665	0.649	0.608	0.736
R1D3			1.000	0.777	0.713	0.906	0.631	0.666	0.676	0.644	0.611	0.724
R1D4				1.000	0.809	0.908	0.610	0.646	0.642	0.678	0.644	0.723
R1D5					1.000	0.875	0.587	0.608	0.611	0.645	0.672	0.701
R1 Total						1.000	0.700	0.734	0.722	0.724	0.700	0.803
R2D1							1.000	0.756	0.769	0.713	0.678	0.877
R2D2								1.000	0.790	0.750	0.688	0.895
R2D3									1.000	0.779	0.709	0.906
R2D4										1.000	0.801	0.907
R2D5											1.000	0.872
R2 Total												1.000

Table 3 provides summary statistics for the distributions of raw scores. The students' scores were rather high, with the average (mean) score being about 88 out of a possible 120. The distributions are quite peaked with approximately 50% of the students having scores between 79 and 97.

Table 3: Grade 11 Writing Summary Statistics

Number of Examinees	129,157	
Total Points	120	
Minimum Score	0	0.0%
25 th Percentile	85.0	70.8%
Median Score	90.0	75.0%
75 th Percentile	97.0	80.8%
Maximum Score	120	100.0%
Mean Score	88.4	73.7%
Variance	239.0	
Standard Deviation	15.5	
Error of Measurement	2.68	
Coefficient Alpha	0.97	

Table 4 and Table 5 give the summary statistics for the domain level data from this analysis. As shown in Table 4, the domain level mean scores are around 6.0 in all cases, which is consistent with the overall writing scores of about 75%. The domain-to-total correlations are consistent with the overall writing scores of about 75%. The domain-to-total correlations are near 0.8, which is to be expected since there are only five domain scores in each total (correlations in Table 4 are calculated as the correlation between the prompt/domain score and the overall writing score summed for all three prompts).

Table 4: Domain Level Means and Correlations

Domain	Narrative		Persuasive		Informational	
	Mean	Corr.	Mean	Corr.	Mean	Corr.
Focus	5.93	0.801	6.02	0.817	6.00	0.792
Content	5.79	0.833	5.93	0.834	5.90	0.821
Organization	5.71	0.826	5.90	0.844	5.80	0.826
Style	5.82	0.839	5.98	0.845	5.87	0.844
Conventions	5.82	0.827	5.94	0.831	5.85	0.824

Table 5 shows the percentage of students scoring in each score category by prompt and domain as well as the cumulative correlations by scoring category. The cumulative correlations were calculated by assigning each scoring category equal to or greater than category X a score of one, and all categories less than X a score of zero, and computing the correlation between this new variable and the students' total raw score.

Table 5: Scoring Category Level Statistics

Prompt	Cat.	Focus		Content		Organization		Style		Conventions	
		%	Corr.	%	Corr.	%	Corr.	%	Corr.	%	Corr.
Narrative	0	2.3	NA	2.3	NA	2.3	NA	2.3	NA	2.3	NA
	2	0.8	0.400	2.0	0.406	2.7	0.404	2.3	0.404	1.8	0.404
	3	0.9	0.430	2.6	0.505	2.5	0.518	2.2	0.510	1.9	0.501
	4	8.4	0.472	12.7	0.556	10.7	0.568	6.8	0.564	7.3	0.555
	5	10.3	0.600	9.4	0.640	12.9	0.643	9.7	0.651	9.3	0.630
	6	54.0	0.609	48.4	0.612	52.1	0.606	51.6	0.636	51.1	0.617
	7	15.7	0.457	15.0	0.459	12.3	0.416	16.5	0.449	18.4	0.449
	8	7.5	0.324	7.6	0.337	4.5	0.264	8.5	0.310	7.9	0.307
Persuasive	0	1.4	NA	1.4	NA	1.4	NA	1.4	NA	1.4	NA
	2	1.7	0.409	3.8	0.412	3.6	0.411	3.4	0.412	2.6	0.411
	3	2.2	0.470	2.4	0.519	2.4	0.526	2.5	0.524	2.2	0.515
	4	8.5	0.523	10.2	0.567	9.1	0.574	7.3	0.577	8.1	0.567
	5	10.2	0.598	11.1	0.628	9.6	0.637	9.5	0.645	9.3	0.619
	6	44.0	0.594	39.8	0.603	47.4	0.621	41.5	0.631	43.4	0.603
	7	19.0	0.485	18.1	0.482	17.3	0.471	19.8	0.484	20.8	0.481
	8	13.0	0.364	13.3	0.374	9.2	0.336	14.8	0.368	12.2	0.346
Informational	0	2.2	NA	2.2	NA	2.2	NA	2.2	NA	2.2	NA
	2	1.7	0.428	2.4	0.431	3.2	0.430	2.5	0.431	2.1	0.432
	3	2.0	0.477	2.7	0.505	2.5	0.513	2.0	0.514	2.4	0.512
	4	9.9	0.520	13.4	0.549	11.7	0.561	9.1	0.564	9.1	0.563
	5	14.0	0.572	10.7	0.612	13.6	0.628	11.2	0.645	13.4	0.624
	6	43.2	0.556	44.4	0.594	48.6	0.595	48.3	0.626	42.4	0.590
	7	17.0	0.470	15.4	0.472	13.1	0.445	15.9	0.469	19.1	0.468
	8	9.9	0.349	8.8	0.364	5.1	0.297	8.7	0.334	9.5	0.330

CALIBRATION AND SCALING

The previous administrations of the grade 11 writing test were calibrated and scaled with a proprietary version of Rasch measurement software developed at DRC. In 2002, the decision was made to move the calibration and scaling of the grade 11 writing test to publicly available software. The program selected for calibration was WINSTEPS¹. This software uses the family of Rasch two-facet measurement models as the basis for a variety of Rasch calibration, equating and scaling functions. WINSTEPS is the industry standard for implementing Rasch measurement models.

Because of the shift in calibration programs and the design of the linking between the 2002 and 2003 administrations, it was necessary to design a multileveled approach to the equating and scaling of the 2003 results. The first step was to recalibrate the three prompts from the 2002 administration. This step was necessary to ensure that the logit step and item values used in linking the two forms were equivalent. Sometimes differences in estimation procedures and different implementation of convergence rules during estimation result in estimates of item and person parameters that vary slightly. To avoid the possibility that this difference might introduce

¹ Linacre, J.M. (2002). WINSTEPS: Rasch Model Computer Programs. Chicago: WINSTEPS.com

unexpected differences into the equating process, the 2002 data was recalibrated using the WINSTEPS program with the default convergence criteria.

In using the WINSTEPS program, there are several choices that must be made. The first is the selection of the measurement model. In this case, the Rasch partial credit model was chosen. The partial credit model estimates the threshold parameters for each item (prompt domain in this assessment) independently and uses the expression δ_{ij} to represent the threshold difficulty. The $k-1$ thresholds, where k is the number of score categories, replace the item difficulty and global threshold parameters associated with the Rasch rating scale model. There is no item difficulty parameter in the partial credit model.

Some authors suggest that the sum of the threshold difficulties represents the item difficulty, but this characterization ignores the dispersion of the thresholds, which can vary quite widely and still sum to the same value. To facilitate simplification of the output (printed and computer files), WINSTEPS reports an item difficulty for all items in a partial credit model and adjusts the threshold difficulties to sum to zero. This is a mere notational convention and does not imply that an item difficulty was estimated for each item. The estimation algorithm in WINSTEPS actually estimates the conventional partial credit model.

The sample of students for the calibration was all students, regardless of norm status, who were classified as attempts on the total score. Students with zero scores on one, two, or all three prompts could be included if no reason was given for exclusion.

The calibrations were done at the domain level, which considers each domain score to be an independent item. With three prompts and five domains, this leads to a fifteen-item test with each item scored zero to eight (zero to four scores from each of two readers). The scores from two readers and 15 items were added together, resulting in a maximum possible score of 120 points.

Because zero scores can only result from an unscorable paper, an unscorable paper is unscorable for all five domains. A paper that receives a zero on any domain will receive zeros on all domains. This lack of independence is a violation of an underlying model assumption of item independence.

Because scaled scores are required for the zero papers, they were included in the final calibrations. However, the reliabilities may be overestimated accordingly.

Equating

Equating writing prompts is always problematic. First, the psychometric behavior of polychotomous items is not as well understood as that of dichotomous items. The actual scoring of writing involves human readers. To equate from year-to-year requires the assumption that the readers in the current year are indistinguishable from those of previous year. It also assumes that the training and control are identical. While everything possible is done to ensure that this is in fact the case, there is some evidence that suggests that readers and trainers do in fact vary.

Between-year equating requires at least one prompt that is used in both administrations. The Persuasive prompt from the 2003 administration provided the link to the 2004 administration. During the recalibration of the three prompts from 2002 for the grade 11 administration, two WINSTEPS output files, an item file and a step file, were generated that were used to anchor the

2003 PSSA Grade Eleven Writing

2003 calibration. A WINSTEPS anchored calibration was run on the 2003 data using the modified item and step files from the 2002 calibration of the linking prompt. The item parameters for the additional two prompts of the 2003 administration were free to vary. Anchoring the item parameters in the 2003 calibration forced the origin of the scale back to the origin established in the calibration of the 2002 administration.

The results of this calibration are shown in Table 6 below.

Table 6: WINSTEPS Calibration for 2004 Grade 11 Writing

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT		OUTFIT		PTMEA CORR.	DISPLACE	Domai	G
					MNSQ	ZSTD	MNSQ	ZSTD				
1	641636	129885	-.27A	.00	.82	-9.9	.74	-9.9	.78	.15	P1D1	0
2	623095	129885	.01A	.00	.75	-9.9	.70	-9.9	.82	.00	P1D2	0
3	611983	129885	.13A	.00	.75	-9.9	.69	-9.9	.80	.01	P1D3	0
4	626849	129885	.01A	.00	.70	-9.9	.64	-9.9	.81	-.02	P1D4	0
5	626955	129885	.05A	.00	.75	-9.9	.69	-9.9	.81	-.08	P1D5	0
6	650197	129559	-.30	.00	.89	-9.9	.86	-9.9	.80	.00	P2D1	0
7	638273	129559	-.16	.00	.86	-9.9	.83	-9.9	.82	.00	P2D2	0
8	633506	129559	-.09	.00	.81	-9.9	.76	-9.9	.82	.00	P2D3	0
9	644716	129559	-.21	.00	.81	-9.9	.78	-9.9	.83	.00	P2D4	0
10	639314	129559	-.14	.00	.87	-9.9	.85	-9.9	.81	.00	P2D5	0
11	645813	128944	-.17	.00	1.02	5.4	1.00	.0	.77	.00	P3D1	0
12	632147	128944	-.04	.00	.93	-9.9	.90	-9.9	.81	.00	P3D2	0
13	619606	128944	.11	.00	.89	-9.9	.85	-9.9	.80	.00	P3D3	0
14	628004	128944	.02	.00	.81	-9.9	.78	-9.9	.82	.00	P3D4	0
15	625521	128944	.03	.00	.90	-9.9	.90	-9.9	.81	.00	P3D5	0
MEAN	632508.	129463.	-.07	.00	.84	-8.9	.80	-9.2				
S.D.	10362.	390.	.13	.00	.08	3.8	.09	2.5				

The numbers listed in the MEASURE column of the above table are the numbers from the 2004 anchored WINSTEPS calibration. The “A” next to each value for prompt one (entry numbers 1 – 5) signifies that these values were *anchored*. WINSTEPS fixed these items at the anchored value (determined by the calibration of data from the 2003 administration of the prompt). In addition, WINSTEPS calculates what the value would have been if the item difficulty had been determined by the 2004 data. The difference between those two values is listed in the DISPLACE column. The value of this column would be added to the value in the measure column to obtain the new value. Positive values in this column indicate that the item is more difficult in this administration than in the last (anchor) administration and negative values indicate that the item is easier than in the last administration.

Numerically the calibrations proceeded without incident. Convergence was achieved in 30 major iterations using the standard Rasch unconditional maximum likelihood procedure. Table 7 gives the logit threshold difficulties for each prompt and domain.

Table 7: Step Logit Difficulties

Prompt	Threshold	Step Logit Difficulties				
		Focus	Content	Organization	Style	Conventions
Narrative	2	-2.10	-3.10	-3.17	-3.09	-3.10
	3	-2.36	-1.56	-1.55	-1.67	-1.75
	4	-3.46	-2.77	-2.60	-2.53	-2.41
	5	-0.44	0.41	0.12	0.02	0.10
	6	-0.41	0.09	0.12	0.04	0.20
	7	3.16	3.31	3.68	3.40	3.31
	8	3.71	3.69	4.30	3.91	3.99
Persuasive	2	-2.60	-3.19	-3.29	-3.23	-3.19
	3	-2.41	-1.79	-1.71	-1.84	-1.91
	4	-2.99	-2.69	-2.46	-2.47	-2.44
	5	-0.64	-0.11	-0.45	-0.46	-0.36
	6	-0.39	-0.15	-0.31	-0.24	-0.11
	7	3.21	3.23	3.53	3.13	3.08
	8	3.73	3.59	4.07	3.65	3.95
Informational	2	-1.90	-2.31	-2.41	-2.33	-2.33
	3	-2.21	-1.80	-1.88	-1.93	-1.97
	4	-3.07	-3.04	-2.65	-2.69	-2.74
	5	-0.73	0.10	-0.27	-0.36	-0.13
	6	-0.21	-0.21	-0.13	-0.10	0.18
	7	3.07	3.29	3.66	3.44	3.13
	8	3.85	3.69	4.45	4.10	4.07

Threshold difficulties can be interpreted in a manner similar to the interpretation of item difficulties for dichotomously scored items. Each threshold is viewed as its own dichotomous item. The logit for threshold 1 is the difficulty of obtaining a score of one, given the person score zero or one. Similarly, the threshold 4 logit is the difficulty of obtaining a four, given the person has obtained a three or four. This formulation does not require the thresholds for a given item to be ordered, although they typically are with writing scores.

The non-independence introduced by the nonscorable papers will partially constrain the threshold difficulties. The other steps are less constrained, but are still quite similar within a prompt. This suggests that the rating scale model would probably fit quite well, although there would be no particular advantage to its use. This probably reflects the consistency with which scores are assigned across domains.

Scaled Scores

Determining student-level scaled scores is a two-step process. First, the raw scores are translated into logits through the use of the WINSTEPS raw-score conversion tables. Second, a linear transformation is applied to the logits to calculate the scaled score.

2003 PSSA Grade Eleven Writing

The grade eleven norm-reference group was originally set-up in 2001 to have a mean scaled score of 1300 and a standard deviation of 100. The linear transformation applied to calculate the scaled score for grade eleven is:

$$\text{Scaled Score} = (110.99 * \text{logit}) + 1132.9 . \quad (1)$$

Table 8 shows the raw score to logit conversion calculated by WINSTEPS. The “E” after the logits corresponding to scores 0 and 120 indicate that these are estimated based on partial scores, as the estimates for zeros and perfect scores in the Rasch model are negative and positive infinity, respectively. In WINSTEPS, this partial score is a variable that can range between 0.01 and 0.99. In this calibration, it was set at the default value of 0.30. As a result, the logit reported for a raw score of 0 was estimated for a raw score of 0.3, and the logit listed for a raw score of 120 was estimated for a raw score of 119.7. The lowest scaled score reported is 700

2003 PSSA Grade Eleven Writing

Table 8: Raw Score to Logit Conversion Table

GRADE 11 WRITING								
Raw Score	Scaled Score	Percentile Rank	Raw Score	Scaled Score	Percentile Rank	Raw Score	Scaled Score	Percentile Rank
0	700	1	46	907	4	92	1348	58
1	700	1	47	912	4	93	1365	62
2	700	1	48	918	5	94	1380	65
3	700	1	49	923	5	95	1395	68
4	700	1	50	930	5	96	1410	71
5	700	1	51	935	5	97	1423	74
6	700	1	52	942	6	98	1438	75
7	700	1	53	948	6	99	1451	78
8	700	1	54	954	6	100	1463	80
9	700	1	55	961	6	101	1477	82
10	700	1	56	969	7	102	1489	84
11	715	1	57	975	7	103	1501	86
12	715	1	58	983	7	104	1513	88
13	731	1	59	991	8	105	1526	89
14	731	1	60	999	9	106	1538	91
15	746	1	61	1006	10	107	1550	92
16	746	1	62	1014	10	108	1563	93
17	759	1	63	1023	11	109	1577	94
18	759	1	64	1032	12	110	1591	95
19	770	1	65	1041	13	111	1605	96
20	770	1	66	1050	13	112	1621	99
21	780	1	67	1059	14	113	1639	99
22	780	1	68	1067	14	114	1659	99
23	789	1	69	1076	15	115	1681	99
24	789	1	70	1085	16	116	1707	99
25	798	1	71	1095	16	117	1741	99
26	798	1	72	1104	17	118	1787	99
27	806	1	73	1113	18	119	1865	99
28	806	1	74	1123	18	120	1999	99
29	813	1	75	1132	20			
30	813	1	76	1142	21			
31	820	1	77	1152	23			
32	827	2	78	1162	24			
33	833	2	79	1172	25			
34	840	2	80	1182	25			
35	846	2	81	1193	25			
36	851	2	82	1204	25			
37	857	3	83	1216	25			
38	863	3	84	1228	29			
39	869	3	85	1240	30			
40	873	3	86	1255	33			
41	879	3	87	1269	37			
42	884	3	88	1284	42			
43	890	3	89	1299	46			
44	896	4	90	1316	52			
45	901	4	91	1331	55			

2003 PSSA Grade Eleven Writing

Once the logits corresponding to each raw score have been determined, the linear transformation described by equation (1) above was applied to the logits. The results are summarized in Table 9 below.

Table 9: Raw Score to Scaled Score Conversion Table

GRADE 11 WRITING								
Raw Score	Scaled Score	Percentile Rank	Raw Score	Scaled Score	Percentile Rank	Raw Score	Scaled Score	Percentile Rank
0	700	1	51	932	3	93	1370	66
10	700	1	52	939	3	94	1388	69
11	719	1	53	944	3	95	1405	72
12	719	1	54	951	4	96	1420	74
13	736	1	55	956	4	97	1435	77
14	736	1	56	963	4	98	1448	79
15	750	1	57	971	5	99	1463	81
16	750	1	58	978	5	100	1475	83
17	762	1	59	984	5	101	1488	85
18	762	1	60	992	6	102	1500	87
19	772	1	61	1000	6	103	1512	88
20	772	1	62	1007	7	104	1524	90
21	782	1	63	1016	7	105	1536	91
22	782	1	64	1024	8	106	1548	92
23	791	1	65	1033	9	107	1560	93
24	791	1	66	1041	10	108	1574	93
25	800	1	67	1050	10	109	1587	95
26	800	1	68	1059	11	110	1600	96
27	807	1	69	1067	12	111	1615	96
28	807	1	70	1076	13	112	1630	97
29	814	1	71	1085	14	113	1648	98
30	814	1	72	1094	15	114	1667	98
31	821	1	73	1103	16	115	1689	99
32	828	1	74	1113	18	116	1716	99
33	834	1	75	1122	19	117	1749	99
34	840	1	76	1132	20	118	1794	99
35	845	1	77	1141	22	119	1872	99
36	851	2	78	1151	23	120	2005	99
37	857	2	79	1161	25			
38	862	2	80	1172	27			
39	868	2	81	1183	27			
40	873	2	82	1194	27			
41	879	2	83	1206	28			
42	883	2	84	1219	28			
43	889	2	85	1233	30			
44	894	2	86	1247	34			
45	899	2	87	1264	39			
46	904	2	88	1281	45			
47	910	2	89	1297	50			
48	915	3	90	1316	56			
49	921	3	91	1335	60			
50	926	3	92	1353	63			

STUDENT-LEVEL RESULTS

Raw score reporting was done in the original metric ranging from 0 to 40 for each prompt and from 0 to 120 for total points achieved (based on adding scores from two readers for all three prompts).

If a prompt is unscorable for a student and no reason is given, the student's scores will include zeros for that prompt. If a valid reason is given for not being tested, the student will be reported as though only the prompts the student attempted were given.

Table 10 shows a few basic statistics for the norming group. Student-level distributions will not generally have the same mean as the school level, and standard deviations will be much larger.

Table 10: Summary Statistics for the Norming Group

Count	Mean	Median	Std. Dev
127481	1315.09	1316	180.35

Table 11 provides a summary of the performance standards in both raw score and scaled score metrics. The scaled score standards are fixed values and will not change unless the standards setting is revisited in the future. The raw score standards will change from year to year as the difficulty of the prompts vary.

The table also shows the impact of the standards in the form of the percent of the 2004 students who were classified in each group. Overall, there were about 70% of the grade eleven students in the Proficient or Advanced categories.

Because not all scaled scores occur in every administration, the minimum score assigned to each category is the lowest scaled score that is greater than or equal to the scaled score standards shown in Table 9.

Table 11: PSSA 2004 Grade 11 Writing Performance Levels

Grade 11	Raw Score 2004	Scaled Score	Percent in Category
Advanced	108	1563 and above	6.8%
Proficient	85	1236-1562	67.2%
Basic	71	1088-1235	15.1%
Below Basic		1087 and below	10.9%

More detail about the student-level scaled score distribution is shown in Table 9. This table gives the raw to scaled score conversions that were used in 2004, as well as the percentile rank assigned to each possible score.

INDICATORS OF CONSISTENCY

Criterion-referenced tests are often used to place the examinees into two or more performance classifications. It is then useful to have some indication of how consistent such classifications

are. Livingston (1972)² argued that a measure of consistency should be a function of the reliability of the test and the distance between the performance standard and the mean of the distribution. Livingston's coefficient, which was formulated from the perspective of True Score Theory, is defined by:

$$1. \kappa^2(X,T) = [\sigma^2_T + (\mu_T - C)^2] / [\sigma^2_X + (\mu_X - C)^2] .$$

If the expressions in parenthesis ($\mu - C$) are zero, equation 1 is the ratio of the true score variance to the observed score variance, which is a standard definition of reliability. The numerator is increased by adding the squared difference between the observed score mean and the cut score.

The coefficient is normally computed, from data, by setting the true score mean equal to the observed score mean and multiplying the observed score variance by an estimate of the test reliability coefficient to obtain a value for the true score variance:

$$2. \kappa^2(X,T) = [r \sigma^2_X + (\mu_X - C)^2] / [\sigma^2_X + (\mu_X - C)^2] .$$

If ($\mu_X - C$) is zero, expression 2 will return the value of the reliability; if it is non-zero, the value will be larger than the reliability.

For the 2004 PSSA, values for the Livingston Coefficient of Consistency are given in Table 12 for each performance standard. To illustrate the calculation, the following data was used to calculate the Grade Eleven Basic Coefficient of Consistency.

$$3. \kappa^2(X,T) = [(.96)(239.0) + (88.4 - 72)^2] / [239.0 + (88.4 - 72)^2] = 0.98 .$$

The necessary ingredients can be found in Table 3 and Table 11.

Table 12: Livingston Coefficient of Consistency at Each Performance Standard

Grade	Basic	Proficient	Advanced
11	0.98	0.96	0.98

SCHOOL-LEVEL RESULTS

Originally, the scaled scores were defined so that the mean of each grade for the Norming group would be 1300 with a standard deviation of 100. In this calculation, each school is treated as one observation with no reference to the number of students.

Table 13: School Level Scores

Group	Schools	Enrollment		Scaled Score	
		Mean	StDev	Mean	StDev
Norming	669	188.5	146.8	1297.0	119.3
Non-Norm	30	49.4	55.4	1188.6	276.4
Total	699	182.6	146.8	1292.5	131.1

² Livingston, S.A. 1972. Criterion Referenced Applications of Classical test Theory. *Journal of Educational Measurement* 9: 13-26.

The Table 13 summarizes the school-level results by Norming and Non-Norming groups. The Non-Norming schools tend to be smaller and lower achieving. There is considerable variation in their performance as shown by the large standard deviations for the scaled scores.

DEFINITION OF GRADE ELEVEN WRITING VARIABLE

The numbers attached to any measurement scale have no intrinsic meaning. Meaning will come with experience. From experience with their own students, educators will understand what a student with a scale score of 1400 is capable of doing. They will know how a student with a score of 1100 is different from a student with a score of 1200.

Few United States residents know how to dress for a temperature expressed on the Celsius scale, but after a few days in Canada or Europe, most learn very quickly how to interpret (or translate) the weather forecast. Along with experience, some mileposts along the way are useful.

1. 0° is freezing.
2. 10° is jacket weather.
3. 20° is room temperature.
4. 30° is hot.
5. 37° is normal body temperature.

Table 14 defines the measurement scale for the PSSA Grade 11 Writing Assessment for scaled scores from 820 to 1730 (Column A). The values presented are intended to give a general understanding of the metric being used; they are not a precise description of any specific administration.

The median score for all students is about 1316, which means 50% of the students are above this point and 50% are below. The table also gives other student percentiles (Column B): the 25th percentile is about 1164 and the 75th about 1415. Percentiles are population-specific statistics and will vary from year to year.

The definition of the performance categories (Column C) is an important component of the PSSA. Written descriptions of what each category means in terms of student behaviors and competencies are available on the Pennsylvania Department of Education website (www.pde.state.pa.us/). In terms of scaled scores, students at or above a score of 1088 will be considered Basic, at or above 1236 is Proficient, and at or above 1563 is Advanced.

The Eleventh Grade Writing Assessment is scored on a four-point rubric in each of five domains (the summation of the two reader's scores result in a total of eight possible points per domain). At the precision of this table, there is little difference among the domains in the difficulties. There are substantial differences between the points on the rubrics and smaller differences among the type of writing (Column D). Any student who makes a reasonable attempt will score at least *two* on all domains. The difficulty associated with achieving a score of *four (limited)* on a domain is about 980 to 1000, depending on the type of writing.

A student at the beginning of the Basic category will have a probability³ of about 90% of scoring at least a *four* on any domain (Column E). A student at the beginning of the proficient category will have a probability over 95% of achieving a *four* and 70% of achieving a *six (adequate)*

³ These probabilities refer to the chances of scoring a *two or higher*.

2003 PSSA Grade Eleven Writing

(Column F). A student at the median score will have a probability of about 80% of achieving a six and a less than 20% chance of achieving an *eight (superior)* (Column G).

Table 14: Grade 11 Writing

Scaled Score	Student Percentile	Performance Category	Prompt Scores	Probability				
				Basic	Prof	Median		
820	10th	Below Basic	Narrative -- Limited Persuasive -- Limited Informational -- Limited	90%	95%	95%		
830								
840								
850								
860								
870								
880								
890								
900								
910								
920								
930								
940								
950								
960								
970								
980								
990								
1000	60%	80%	90%					
1010								
1020								
1030								
1040								
1050								
1060								
1070								
1080								
1090	25th	Basic		50%	70%	80%		
1100				40%				
1110								
1120								
1130								
1140								
1150		Basic		30%	60%	70%		
1160								
1170								
1180								
1190								
1200								
1210								
1220								
1230								
1240		Proficient		20%	50%			
1250								

2003 PSSA Grade Eleven Writing

Scaled Score	Student Percentile	Performance Category	Prompt Scores	Basic	Probability Prof	Median
1260	Median		Narrative -- Adequate Persuasive -- Adequate Informational -- Adequate	10%	40%	60%
1270						
1280						
1290						
1300						
1310						
1320						
1330						
1340	75th				20%	40%
1350						
1360						
1370						
1380						
1390						
1400						
1410						
1420						
1430						
1440						
1450						
1460						
1470						
1480						
1490						
1500	90th				10%	20%
1510						
1520						
1530						
1540						
1550						
1560						
1560		Proficient				10%
1570	95th	Advanced				
1580						
1590						
1600						
1610						
1620						
1630						
1640						
1650						
1660						
1670						
1680						
1690						
1700						
1710						
1720						
1730						
1730			Persuasive -- Superior Informational -- Superior Narrative -- Superior			

Appendix A: WINSTEPS Output Tables

This appendix to the 2004 Grade 11 Writing Assessment Technical Analysis provides WINSTEPS output tables generated during the calibration, equating and analysis of the data.

ITEM CALIBRATION

The numbers listed in the MEASURE column of Table 1 below are the item difficulties from the 2004 anchored WINSTEPS calibration. The “A” next to each value for prompt one (entry numbers 1 – 5) signifies that these values were *anchored*. WINSTEPS fixed these items at the anchored value (determined by the calibration of data from the 2003 administrations of the prompt). In addition, WINSTEPS calculates what the value would have been if the item difficulty had been determined by the 2004 data. The difference between those two values is listed in the DISPLACE column. The value of this column would be added to the value in the measure column to obtain the new value. Positive values in this column indicate that the item is more difficult in this administration than in the previous (anchor) administration and negative values indicate that the item is easier than in the previous administration.

Table 1: Grade 11 Item Calibrations

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEA CORR.	DISPLACE	Domain	G
1	641636	129885	-.27A	.00	.82	-9.9	.74	-9.9	.78	.15	P1D1	0
2	623095	129885	.01A	.00	.75	-9.9	.70	-9.9	.82	.00	P1D2	0
3	611983	129885	.13A	.00	.75	-9.9	.69	-9.9	.80	.01	P1D3	0
4	626849	129885	.01A	.00	.70	-9.9	.64	-9.9	.81	-.02	P1D4	0
5	626955	129885	.05A	.00	.75	-9.9	.69	-9.9	.81	-.08	P1D5	0
6	650197	129559	-.30	.00	.89	-9.9	.86	-9.9	.80	.00	P2D1	0
7	638273	129559	-.16	.00	.86	-9.9	.83	-9.9	.82	.00	P2D2	0
8	633506	129559	-.09	.00	.81	-9.9	.76	-9.9	.82	.00	P2D3	0
9	644716	129559	-.21	.00	.81	-9.9	.78	-9.9	.83	.00	P2D4	0
10	639314	129559	-.14	.00	.87	-9.9	.85	-9.9	.81	.00	P2D5	0
11	645813	128944	-.17	.00	1.02	5.4	1.00	.0	.77	.00	P3D1	0
12	632147	128944	-.04	.00	.93	-9.9	.90	-9.9	.81	.00	P3D2	0
13	619606	128944	.11	.00	.89	-9.9	.85	-9.9	.80	.00	P3D3	0
14	628004	128944	.02	.00	.81	-9.9	.78	-9.9	.82	.00	P3D4	0
15	625521	128944	.03	.00	.90	-9.9	.90	-9.9	.81	.00	P3D5	0
MEAN	632508.	129463.	-.07	.00	.84	-8.9	.80	-9.2				
S.D.	10362.	390.	.13	.00	.08	3.8	.09	2.5				

RAW SCORE/LOGIT CONVERSION TABLES

The following tables summarize the raw score to logit conversion tables (which are the first step in producing the raw score to scaled score conversion tables) produced by WINSTEPS for the 2004 Grade 11 Writing Assessment data. Table 5 presents the raw scores for the entire test. Table 6 through Table 8 summarize the raw score to logit conversions by prompt. Table 9 through Table 13 summarize the raw score to logit conversions by domain for each prompt.

It is a requirement of the WINSTEPS calibration procedure that the scoring categories for a data set be ordered. For the Grade 11 Writing Assessment, possible scores on an item (prompt-domain) range from zero to eight points. However, a student cannot receive a score of one (due to the fact that the scoring requirements do not allow one reader to assign a one while the other reader assigns a zero. In order to perform the calibration, WINSTEPS recodes the scoring categories as summarized in Tables 2 through 4.

Table 2: WINSTEPS Category to Raw Score Conversion – Domain Level

WINSTEPS Category	Raw Score
0	0
1	2
2	3
3	4
4	5
5	6
6	7
7	8

Table 3: WINSTEPS Category to Raw Score Conversion – Prompt Level

WINSTEPS Category	Raw Score	WINSTEPS Category	Raw Score
0	0	18	23
1	NA	19	24
2	NA	20	25
3	NA	21	26
4	NA	22	27
5	10	23	28
6	11	24	29
7	12	25	30
8	13	26	31
9	14	27	32
10	15	28	33
11	16	29	34
12	17	30	35
13	18	31	36
14	19	32	37
15	20	33	38
16	21	34	39
17	22	35	40

Table 4: WINSTEPS Category to Raw Score Conversion – Total Test

WINSTEPS Category	Raw Score	WINSTEPS Category	Raw Score	WINSTEPS Category	Raw Score
0	0	30	45	69	84
1	NA	31	46	70	85
2	NA	32	47	71	86
3	NA	33	48	72	87
4	NA	34	49	73	88
5	10	35	50	74	89
6	11	36	51	75	90
6	12	37	52	76	91
7	13	38	53	77	92
7	14	39	54	78	93
8	15	40	55	79	94
8	16	41	56	80	95
9	17	42	57	81	96
9	18	43	58	82	97
10	19	44	59	83	98
10	20	45	60	84	99
11	21	46	61	85	100
11	22	47	62	86	101
12	23	48	63	87	102
12	24	49	64	88	103
13	25	50	65	89	104
13	26	51	66	90	105
14	27	52	67	91	106
14	28	53	68	92	107
15	29	54	69	93	108
15	30	55	70	94	109
16	31	56	71	95	110
17	32	57	72	96	111
18	33	58	73	97	112
19	34	59	74	98	113
20	35	60	75	99	114
21	36	61	76	100	115
22	37	62	77	101	116
23	38	63	78	102	117
24	39	64	79	103	118
25	40	65	80	104	119
26	41	66	81	105	120
27	42	67	82		
28	43	68	83		

When looking at the conversion shown in Table 4 (total test), it should become apparent that the WINSTEPS scoring categories 6 through 15 each correspond to two possible raw scores. Since the data was calibrated using non-contiguous scoring categories (0, 2, 3, 4, 5, 6, 7, 8),

WINSTEPS only calculates *seven* step logits per prompt-domain. As a result of this issue, there exists only $15 \times 7 = 105$ logits to apply to $15 \times 8 = 120$ raw scores. It was determined that the data best fit the model by removing every other raw score point on the very bottom of the scoring range – leading to situation displayed in Table 4.

Table 5: Grade 11 Raw Score/Logit Conversion Table – Total Test

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-6.76E	1.83	36	-1.81	.23	72	1.18	.38
1	-5.54	1.01	37	-1.75	.23	73	1.33	.39
2	-4.83	.72	38	-1.70	.24	74	1.48	.40
3	-4.42	.58	39	-1.64	.24	75	1.65	.41
4	-4.13	.50	40	-1.59	.24	76	1.82	.41
5	-3.90	.44	41	-1.53	.25	77	1.98	.41
6	-3.73	.40	42	-1.46	.25	78	2.14	.40
7	-3.58	.37	43	-1.40	.25	79	2.30	.39
8	-3.45	.34	44	-1.34	.26	80	2.45	.38
9	-3.34	.32	45	-1.27	.26	81	2.59	.37
10	-3.25	.30	46	-1.20	.27	82	2.72	.36
11	-3.16	.29	47	-1.13	.27	83	2.84	.35
12	-3.08	.28	48	-1.05	.27	84	2.97	.34
13	-3.00	.27	49	-.98	.27	85	3.08	.34
14	-2.94	.26	50	-.90	.28	86	3.20	.33
15	-2.87	.25	51	-.83	.28	87	3.31	.33
16	-2.81	.24	52	-.75	.28	88	3.42	.33
17	-2.75	.24	53	-.67	.28	89	3.52	.33
18	-2.69	.23	54	-.59	.28	90	3.63	.33
19	-2.64	.23	55	-.51	.28	91	3.74	.33
20	-2.59	.23	56	-.43	.28	92	3.85	.34
21	-2.54	.22	57	-.35	.29	93	3.97	.34
22	-2.49	.22	58	-.27	.29	94	4.09	.35
23	-2.44	.22	59	-.18	.29	95	4.21	.36
24	-2.39	.22	60	-.10	.29	96	4.34	.37
25	-2.34	.22	61	-.01	.29	97	4.48	.38
26	-2.29	.22	62	.07	.30	98	4.64	.40
27	-2.25	.22	63	.16	.30	99	4.81	.43
28	-2.20	.22	64	.25	.31	100	5.01	.46
29	-2.15	.22	65	.35	.31	101	5.25	.51
30	-2.11	.22	66	.45	.32	102	5.55	.59
31	-2.06	.22	67	.55	.33	103	5.96	.71
32	-2.01	.22	68	.66	.34	104	6.66	1.00
33	-1.96	.22	69	.78	.35	105	7.86E	1.83
34	-1.91	.22	70	.90	.36			
35	-1.86	.23	71	1.03	.37			

Table 6: Grade 11 Raw Score/Logit Conversion Table – Prompt 1 (Narrative)

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-5.76E	1.85	12	-1.75	.42	24	1.30	.64
1	-4.51	1.02	13	-1.57	.44	25	1.75	.69
2	-3.80	.71	14	-1.37	.46	26	2.22	.67
3	-3.40	.57	15	-1.16	.48	27	2.65	.63
4	-3.12	.49	16	-.92	.49	28	3.02	.59
5	-2.90	.45	17	-.67	.50	29	3.35	.57
6	-2.71	.42	18	-.43	.50	30	3.67	.57
7	-2.55	.40	19	-.18	.50	31	4.00	.59
8	-2.39	.39	20	.07	.50	32	4.37	.63
9	-2.24	.39	21	.33	.52	33	4.83	.74
10	-2.08	.40	22	.61	.54	34	5.56	1.01
11	-1.92	.40	23	.93	.59	35	6.76E	1.82

Table 7: Grade 11 Raw Score/Logit Conversion Table – Prompt 2 (Persuasive)

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-5.92E	1.85	12	-1.86	.40	24	1.03	.67
1	-4.67	1.03	13	-1.69	.41	25	1.53	.72
2	-3.94	.72	14	-1.52	.42	26	2.04	.70
3	-3.52	.58	15	-1.33	.44	27	2.49	.64
4	-3.23	.51	16	-1.13	.45	28	2.87	.60
5	-3.00	.46	17	-.93	.46	29	3.21	.57
6	-2.80	.43	18	-.71	.47	30	3.53	.57
7	-2.63	.41	19	-.48	.48	31	3.86	.59
8	-2.47	.40	20	-.24	.49	32	4.23	.63
9	-2.32	.39	21	.01	.52	33	4.69	.74
10	-2.17	.39	22	.29	.55	34	5.42	1.01
11	-2.01	.39	23	.62	.60	35	6.63E	1.82

Table 8: Grade 11 Raw Score/Logit Conversion Table – Prompt 3 (Informational)

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-5.12E	1.79	12	-1.81	.38	24	1.17	.67
1	-4.01	.93	13	-1.66	.40	25	1.65	.72
2	-3.44	.62	14	-1.50	.42	26	2.16	.70
3	-3.13	.50	15	-1.31	.45	27	2.62	.65
4	-2.92	.44	16	-1.10	.47	28	3.01	.61
5	-2.74	.40	17	-.87	.49	29	3.36	.58
6	-2.59	.38	18	-.63	.49	30	3.69	.58
7	-2.46	.36	19	-.39	.50	31	4.04	.60
8	-2.33	.35	20	-.13	.51	32	4.42	.65
9	-2.20	.35	21	.13	.53	33	4.90	.75
10	-2.08	.36	22	.43	.56	34	5.65	1.02
11	-1.95	.36	23	.76	.60	35	6.87E	1.83

Table 9: Grade 11 Raw Score/Logit Conversion Table – Domain 1 (Focus)

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-4.70E	1.70	8	-1.90	.51	16	2.26	.86
1	-3.74	.85	9	-1.61	.56	17	2.91	.76
2	-3.26	.59	10	-1.26	.62	18	3.46	.73
3	-2.97	.50	11	-.86	.64	19	4.03	.79
4	-2.74	.46	12	-.44	.66	20	4.82	1.03
5	-2.54	.44	13	.03	.71	21	6.05E	1.83
6	-2.34	.45	14	.61	.83			
7	-2.14	.47	15	1.42	.95			

Table 10: Grade 11 Raw Score/Logit Conversion Table – Domain 2 (Content)

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-5.20E	1.86	8	-1.52	.57	16	2.38	.81
1	-3.94	1.01	9	-1.17	.62	17	2.98	.73
2	-3.25	.70	10	-.77	.64	18	3.49	.71
3	-2.86	.57	11	-.36	.64	19	4.03	.77
4	-2.57	.52	12	.05	.64	20	4.79	1.01
5	-2.31	.49	13	.47	.68	21	5.98E	1.81
6	-2.07	.50	14	.99	.76			
7	-1.81	.52	15	1.66	.86			

Table 11: Grade 11 Raw Score/Logit Conversion Table – Domain 3 (Organization)

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-5.27E	1.87	8	-1.50	.53	16	2.73	.89
1	-3.98	1.04	9	-1.19	.57	17	3.41	.77
2	-3.24	.72	10	-.85	.60	18	3.97	.74
3	-2.82	.59	11	-.47	.62	19	4.55	.80
4	-2.52	.53	12	-.07	.65	20	5.35	1.04
5	-2.26	.50	13	.38	.71	21	6.59E	1.83
6	-2.01	.49	14	.96	.83			
7	-1.76	.51	15	1.81	.99			

Table 12: Grade 11 Raw Score/Logit Conversion Table – Domain 4 (Style)

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-5.22E	1.86	8	-1.53	.53	16	2.45	.85
1	-3.96	1.02	9	-1.23	.57	17	3.08	.75
2	-3.25	.71	10	-.89	.60	18	3.62	.73
3	-2.84	.58	11	-.51	.62	19	4.19	.79
4	-2.54	.52	12	-.12	.64	20	4.98	1.03
5	-2.29	.49	13	.33	.70	21	6.21E	1.83
6	-2.05	.49	14	.88	.80			
7	-1.80	.50	15	1.64	.92			

Table 13: Grade 11 Raw Score/Logit Conversion Table – Domain 5 (Conventions)

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
-------	---------	------	-------	---------	------	-------	---------	------

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	-5.21E	1.85	8	-1.49	.55	16	2.43	.83
1	-3.96	1.02	9	-1.16	.60	17	3.06	.76
2	-3.26	.71	10	-.78	.63	18	3.62	.75
3	-2.85	.58	11	-.38	.64	19	4.21	.81
4	-2.55	.52	12	.03	.65	20	5.03	1.05
5	-2.29	.50	13	.47	.69	21	6.29E	1.84
6	-2.04	.50	14	1.00	.78			
7	-1.78	.52	15	1.69	.87			

THRESHOLD DIFFICULTIES

Tables 14 to 28 contain the threshold difficulties for the 15 items (3 prompts x 5 domains) for the 2004 administration. The “A” next to the structure measure in the first panel for each domain, indicates that prompt 1 threshold difficulties were anchored to the 2003 values. The threshold difficulties for the remaining 10 items were free to vary.

Table 14: Grade 11 Threshold Difficulties – Prompt 1 Domain 1

Domain ITEM DIFFICULTY MEASURE OF -.27 ADDED TO MEASURES

CATEGORY		OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	STRUCTURE	CATEGORY		
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	1094	1	-1.83	-1.85	2.13	2.90	NONE A	(-3.90)	0 Unscorable
2	1	295	0	-2.58*	-3.94	.50	.47	-1.83A	-2.99	2 Minimal
3	2	651	0	-2.08	-2.17	.51	.46	-2.09A	-2.40	3
4	3	9383	7	-.81	-.68	.86	.79	-3.19A	-1.65	4 Limited
5	4	14791	11	.28	.50	.84	.79	-.17A	-.40	5
6	5	77351	59	1.61	1.28	.97	.95	-.14A	1.42	6 Adequate
7	6	18269	14	3.08	3.34	.62	.61	3.43A	3.46	7
8	7	8051	6	4.13	6.28	.64	.60	3.98A	(5.10)	8 Superior
MISSING		440	0	.18						

AVERAGE MEASURE is mean of measures in category.

+										
CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		OBS-EXP	
LABEL	MEASURE	S.E.	AT CAT.	----	ZONE----	PROBABLT	M->C	C->M	RESIDUAL	
+										
0	NONE		(-3.90)	-INF	-3.45		53%	4%	45.4%	0 Unscorable
2	-2.10	.04	-2.99	-3.45	-2.68	-2.99	21%	43%	-50.0%	2 Minimal
3	-2.36	.04	-2.40	-2.68	-2.08	-2.63	15%	34%	-35.8%	3
4	-3.46	.03	-1.65	-2.08	-1.04	-2.39	56%	35%	-14.4%	4 Limited
5	-.44	.01	-.40	-1.04	.32	-.85	36%	43%	-5.0%	5
6	-.41	.01	1.42	.32	2.61	.04	80%	74%	18.8%	6 Adequate
7	3.16	.01	3.46	2.61	4.36	2.85	39%	62%	-19.3%	7
8	3.71	.01	(5.10)	4.36	+INF	4.06	68%	35%	-39.2%	8 Superior

Table 15: Grade 11 Threshold Difficulties – Prompt 1 Domain 2

Domain ITEM DIFFICULTY MEASURE OF .01 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	1094	1	-1.83	-1.62	1.82	2.05	NONE A	(-4.37)	0 Unscorable
2	1	1336	1	-2.28*	-2.68	.46	.42	-3.11A	-2.87	2 Minimal
3	2	1695	1	-1.54	-1.37	.45	.43	-1.57A	-2.01	3
4	3	14415	11	-.41	-.19	.67	.61	-2.78A	-1.01	4 Limited
5	4	17073	13	.63	.83	.77	.73	.40A	.26	5
6	5	67824	52	1.76	1.49	.96	.93	.08A	1.78	6 Adequate
7	6	17424	13	3.05	3.10	.68	.69	3.30A	3.53	7
8	7	9024	7	4.07	5.27	.71	.69	3.68A	(5.12)	8 Superior
MISSING	440	0		.17						

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		OBS-EXP	
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----		PROBABLT	M->C	C->M	RESIDUAL	
0	NONE		(-4.37)	-INF	-3.61		56%	3%	48.3%	0 Unscorable
2	-3.10	.04	-2.87	-3.61	-2.40	-3.34	52%	43%	-31.8%	2 Minimal
3	-1.56	.03	-2.01	-2.40	-1.57	-2.24	24%	43%	-21.0%	3
4	-2.77	.02	-1.01	-1.57	-.35	-1.87	60%	42%	-21.3%	4 Limited
5	.41	.01	.26	-.35	.90	-.12	34%	48%	-.2%	5
6	.09	.01	1.78	.90	2.75	.63	76%	70%	18.6%	6 Adequate
7	3.31	.01	3.53	2.75	4.40	2.97	38%	56%	-11.8%	7
8	3.69	.01	(5.12)	4.40	+INF	4.08	71%	32%	-28.6%	8 Superior

Table 16: Grade 11 Threshold Difficulties – Prompt 1 Domain 3

Domain ITEM DIFFICULTY MEASURE OF .13 ADDED TO MEASURES

CATEGORY		OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY		
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	1094	1	-1.83	-1.55	1.83	1.99	NONE A	(-4.42)	0 Unscorable
2	1	1630	1	-2.23*	-2.38	.48	.44	-3.30A	-2.85	2 Minimal
3	2	1894	1	-1.48	-1.34	.47	.42	-1.68A	-1.97	3
4	3	12790	10	-.43	-.26	.67	.59	-2.73A	-1.04	4 Limited
5	4	18215	14	.66	.86	.82	.78	-.01A	.17	5
6	5	74835	57	1.86	1.61	.97	.96	-.01A	1.95	6 Adequate
7	6	14829	11	3.32	3.46	.66	.68	3.55A	4.01	7
8	7	4598	4	4.32	6.56	.75	.69	4.17A	(5.68)	8 Superior
MISSING		440	0	.17						

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		OBS-EXP	
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----		PROBABLT	M->C	C->M	RESIDUAL	
0	NONE		(-4.42)	-INF	-3.64		56%	3%	49.7%	0 Unscorable
2	-3.17	.04	-2.85	-3.64	-2.37	-3.38	57%	43%	-22.3%	2 Minimal
3	-1.55	.02	-1.97	-2.37	-1.55	-2.20	26%	40%	-16.3%	3
4	-2.60	.02	-1.04	-1.55	-.45	-1.80	57%	41%	-19.0%	4 Limited
5	.12	.01	.17	-.45	.87	-.28	36%	49%	-5.1%	5
6	.12	.01	1.95	.87	3.15	.58	79%	74%	16.0%	6 Adequate
7	3.68	.01	4.01	3.15	4.93	3.38	41%	54%	-16.0%	7
8	4.30	.01	(5.68)	4.93	+INF	4.63	59%	25%	-39.8%	8 Superior

Table 17: Grade 11 Threshold Difficulties – Prompt 1 Domain 4

Domain ITEM DIFFICULTY MEASURE OF .01 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	1094	1	-1.83	-1.63	1.86	2.02	NONE A	(-4.37)	0 Unscorable
2	1	1412	1	-2.30*	-2.55	.42	.38	-3.10A	-2.85	2 Minimal
3	2	1825	1	-1.57	-1.44	.42	.37	-1.68A	-1.99	3
4	3	10618	8	-.59	-.33	.56	.46	-2.54A	-1.08	4 Limited
5	4	16988	13	.44	.81	.71	.64	.01A	.08	5
6	5	72845	56	1.75	1.48	.92	.90	.03A	1.77	6 Adequate
7	6	17965	14	3.11	3.06	.69	.69	3.39A	3.68	7
8	7	7138	5	4.16	5.79	.74	.70	3.90A	(5.31)	8 Superior
MISSING	440	0		.17						

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		OBS-EXP	
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----		PROBABLT	M->C	C->M	RESIDUAL	
0	NONE		(-4.37)	-INF	-3.61		56%	3%	46.5%	0 Unscorable
2	-3.09	.04	-2.85	-3.61	-2.38	-3.33	53%	47%	-27.9%	2 Minimal
3	-1.67	.03	-1.99	-2.38	-1.57	-2.23	27%	43%	-21.5%	3
4	-2.53	.02	-1.08	-1.57	-.51	-1.80	55%	43%	-26.7%	4 Limited
5	.02	.01	.08	-.51	.76	-.36	39%	53%	-7.2%	5
6	.04	.01	1.77	.76	2.86	.49	79%	74%	18.4%	6 Adequate
7	3.40	.01	3.68	2.86	4.57	3.09	42%	55%	-9.1%	7
8	3.91	.01	(5.31)	4.57	+INF	4.27	68%	31%	-33.8%	8 Superior

Table 18: Grade 11 Threshold Difficulties – Prompt 1 Domain 5

Domain ITEM DIFFICULTY MEASURE OF .05 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	1094	1	-1.83	-1.61	1.82	1.96	NONE A	(-4.38)	0 Unscorable
2	1	1270	1	-2.31*	-2.89	.40	.36	-3.15A	-2.86	2 Minimal
3	2	1772	1	-1.57	-1.66	.42	.38	-1.80A	-1.96	3
4	3	11135	9	-.50	-.27	.64	.56	-2.46A	-1.01	4 Limited
5	4	17116	13	.47	.91	.80	.71	.05A	.19	5
6	5	71871	55	1.73	1.45	1.02	.97	.15A	1.81	6 Adequate
7	6	18472	14	3.11	3.13	.68	.70	3.26A	3.68	7
8	7	7155	5	4.13	5.62	.76	.73	3.94A	(5.35)	8 Superior
MISSING	440	0		.17						

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		OBS-EXP	
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----	PROBABLT	M->C	C->M	RESIDUAL		
0	NONE		(-4.38)	-INF	-3.63		56%	3%	47.0%	0 Unscorable
2	-3.10	.04	-2.86	-3.63	-2.38	-3.35	49%	48%	-36.6%	2 Minimal
3	-1.75	.02	-1.96	-2.38	-1.53	-2.24	25%	42%	-33.6%	3
4	-2.41	.02	-1.01	-1.53	-.42	-1.76	51%	46%	-28.9%	4 Limited
5	.10	.01	.19	-.42	.87	-.26	36%	50%	-12.0%	5
6	.20	.01	1.81	.87	2.85	.62	78%	70%	23.7%	6 Adequate
7	3.31	.01	3.68	2.85	4.60	3.04	42%	53%	-11.0%	7
8	3.99	.01	(5.35)	4.60	+INF	4.31	66%	30%	-31.9%	8 Superior

Table 19: Grade 11 Threshold Difficulties – Prompt 2 Domain 1

Domain ITEM DIFFICULTY MEASURE OF -.30 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	602	0	-2.30	-2.59	1.42	1.68	NONE	(-4.17)	0 Unscorable
2	1	774	1	-2.39*	-2.11	.63	.59	-2.30	-3.07	2 Minimal
3	2	1402	1	-1.76	-1.49	.66	.60	-2.11	-2.39	3
4	3	9462	7	-.60	-.63	.99	.97	-2.69	-1.61	4 Limited
5	4	16216	12	.38	.45	.89	.87	-.34	-.45	5
6	5	66116	51	1.52	1.57	.87	.85	-.09	1.43	6 Adequate
7	6	22120	17	2.82	2.66	.80	.77	3.51	3.49	7
8	7	12867	10	3.81	3.74	.90	.83	4.03	(5.13)	8 Superior
MISSING		766	1	.19						

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----		PROBABLT	M->C	C->M	
0	NONE		(-4.17)	-INF	-3.61		56%	6%	0 Unscorable
2	-2.60	.05	-3.07	-3.61	-2.71	-3.21	44%	29%	2 Minimal
3	-2.41	.03	-2.39	-2.71	-2.04	-2.65	27%	31%	3
4	-2.99	.02	-1.61	-2.04	-1.06	-2.27	51%	27%	4 Limited
5	-.64	.01	-.45	-1.06	.28	-.95	35%	40%	5
6	-.39	.01	1.43	.28	2.65	.01	71%	77%	6 Adequate
7	3.21	.01	3.49	2.65	4.39	2.89	41%	53%	7
8	3.73	.01	(5.13)	4.39	+INF	4.08	75%	24%	8 Superior

Table 20: Grade 11 Threshold Difficulties – Prompt 2 Domain 2

Domain ITEM DIFFICULTY MEASURE OF -.16 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	602	0	-2.30	-2.49	1.26	1.36	NONE	(-4.48)	0 Unscorable
2	1	1625	1	-2.23	-1.93	.58	.52	-3.03	-2.98	2 Minimal
3	2	1981	2	-1.47	-1.24	.61	.55	-1.63	-2.12	3
4	3	12971	10	-.38	-.36	.87	.82	-2.53	-1.23	4 Limited
5	4	16631	13	.58	.65	.84	.83	.05	-.08	5
6	5	61094	47	1.62	1.65	.89	.86	.01	1.58	6 Adequate
7	6	20806	16	2.80	2.66	.83	.80	3.39	3.44	7
8	7	13849	11	3.78	3.71	.91	.86	3.75	(5.02)	8 Superior
MISSING		766	1	.19						

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----		PROBABLT	M->C	C->M	
0	NONE		(-4.48)	-INF	-3.72		54%	4%	0 Unscorable
2	-3.19	.05	-2.98	-3.72	-2.51	-3.44	58%	32%	2 Minimal
3	-1.79	.03	-2.12	-2.51	-1.71	-2.36	27%	32%	3
4	-2.69	.02	-1.23	-1.71	-.67	-1.95	55%	34%	4 Limited
5	-.11	.01	-.08	-.67	.59	-.51	33%	41%	5
6	-.15	.01	1.58	.59	2.64	.32	69%	73%	6 Adequate
7	3.23	.01	3.44	2.64	4.30	2.88	38%	51%	7
8	3.59	.01	(5.02)	4.30	+INF	3.98	75%	27%	8 Superior

Table 21: Grade 11 Threshold Difficulties – Prompt 2 Domain 3

Domain ITEM DIFFICULTY MEASURE OF -.09 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	602	0	-2.30	-2.47	1.27	1.37	NONE	(-4.54)	0 Unscorable
2	1	1816	1	-2.20	-1.93	.56	.51	-3.20	-2.94	2 Minimal
3	2	2026	2	-1.47	-1.28	.63	.53	-1.62	-2.06	3
4	3	9920	8	-.52	-.45	.80	.70	-2.37	-1.28	4 Limited
5	4	16470	13	.45	.58	.79	.76	-.36	-.27	5
6	5	69821	54	1.66	1.68	.84	.82	-.22	1.64	6 Adequate
7	6	19435	15	3.02	2.80	.75	.73	3.62	3.81	7
8	7	9469	7	4.00	3.90	.89	.80	4.16	(5.46)	8 Superior
MISSING	766	1	.19							

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----		PROBABLT	M->C	C->M	
0	NONE		(-4.54)	-INF	-3.75		54%	4%	0 Unscorable
2	-3.29	.05	-2.94	-3.75	-2.46	-3.50	60%	33%	2 Minimal
3	-1.72	.03	-2.06	-2.46	-1.69	-2.29	27%	33%	3
4	-2.47	.02	-1.28	-1.69	-.81	-1.85	50%	32%	4 Limited
5	-.45	.01	-.27	-.81	.42	-.74	37%	40%	5
6	-.31	.01	1.64	.42	2.95	.13	74%	80%	6 Adequate
7	3.52	.01	3.81	2.95	4.72	3.20	42%	51%	7
8	4.06	.01	(5.46)	4.72	+INF	4.41	74%	20%	8 Superior

Table 22: Grade 11 Threshold Difficulties – Prompt 2 Domain 4

Domain ITEM DIFFICULTY MEASURE OF -.21 ADDED TO MEASURES

CATEGORY		OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE		CATEGORY	
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	

0	0	602	0	-2.30	-2.47	1.28	1.38	NONE	(-4.51)	0 Unscorable
2	1	1681	1	-2.26	-1.94	.51	.45	-3.02	-2.96	2 Minimal
3	2	2083	2	-1.52	-1.28	.57	.47	-1.63	-2.10	3
4	3	10165	8	-.54	-.45	.75	.66	-2.26	-1.29	4 Limited
5	4	16704	13	.42	.56	.74	.73	-.25	-.25	5
6	5	62124	48	1.60	1.59	.83	.81	-.03	1.48	6 Adequate
7	6	22462	17	2.77	2.64	.85	.82	3.34	3.42	7
8	7	13738	11	3.75	3.71	.94	.89	3.86	(5.05)	8 Superior

MISSING		766	1	.19						

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----		PROBABLT	M->C	C->M	
0	NONE		(-4.51)	-INF	-3.74		54%	4%	0 Unscorable
2	-3.23	.05	-2.96	-3.74	-2.49	-3.47	60%	32%	2 Minimal
3	-1.83	.03	-2.10	-2.49	-1.71	-2.34	28%	32%	3
4	-2.47	.02	-1.29	-1.71	-.80	-1.88	51%	33%	4 Limited
5	-.45	.01	-.25	-.80	.43	-.72	39%	41%	5
6	-.24	.01	1.48	.43	2.59	.17	70%	78%	6 Adequate
7	3.13	.01	3.42	2.59	4.31	2.82	40%	50%	7
8	3.65	.01	(5.05)	4.31	+INF	4.00	73%	26%	8 Superior

Table 23: Grade 11 Threshold Difficulties – Prompt 2 Domain 5

Domain ITEM DIFFICULTY MEASURE OF -.14 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	602	0	-2.30	-2.49	1.27	1.36	NONE	(-4.49)	0 Unscorable
2	1	1633	1	-2.20	-1.94	.59	.54	-3.05	-2.97	2 Minimal
3	2	2231	2	-1.43	-1.26	.68	.60	-1.77	-2.10	3
4	3	11079	9	-.39	-.40	.90	.87	-2.30	-1.25	4 Limited
5	4	17592	13	.51	.61	.85	.82	-.22	-.16	5
6	5	60577	46	1.61	1.63	.91	.89	.03	1.53	6 Adequate
7	6	24186	19	2.82	2.68	.82	.83	3.22	3.54	7
8	7	11659	9	3.81	3.76	.94	.90	4.09	(5.28)	8 Superior
MISSING	766	1	.19							

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		OBS-EXP	
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----		PROBABLT	M->C	C->M	RESIDUAL	
0	NONE		(-4.49)	-INF	-3.73		54%	4%	-1.4%	0 Unscorable
2	-3.19	.05	-2.97	-3.73	-2.50	-3.45	57%	31%	-1.0%	2 Minimal
3	-1.91	.03	-2.10	-2.50	-1.70	-2.36	28%	30%	-.6%	3
4	-2.44	.02	-1.25	-1.70	-.73	-1.87	48%	32%	-.3%	4 Limited
5	-.36	.01	-.16	-.73	.52	-.64	36%	40%	-.1%	5
6	-.11	.01	1.53	.52	2.66	.28	67%	74%	.1%	6 Adequate
7	3.08	.01	3.54	2.66	4.50	2.84	43%	53%	.1%	7
8	3.95	.01	(5.28)	4.50	+INF	4.23	72%	20%	.2%	8 Superior

Table 24: Grade 11 Threshold Difficulties – Prompt 3 Domain 1

Domain ITEM DIFFICULTY MEASURE OF -.17 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	1001	1	-2.11	-2.48	1.68	2.20	NONE	(-3.69)	0 Unscorable
2	1	717	1	-2.19*	-2.01	.77	.88	-1.73	-2.78	2 Minimal
3	2	1175	1	-1.64	-1.41	.75	.74	-2.04	-2.20	3
4	3	9175	7	-.46	-.58	1.12	1.15	-2.90	-1.54	4 Limited
5	4	17926	14	.52	.48	1.06	1.04	-.56	-.42	5
6	5	62559	48	1.51	1.58	1.02	.97	-.04	1.45	6 Adequate
7	6	24015	18	2.74	2.65	.88	.87	3.24	3.49	7
8	7	12376	9	3.77	3.74	.95	.92	4.02	(5.20)	8 Superior
MISSING	1381	1	.15							

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE	
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----	PROBABLT	M->C	C->M	
0	NONE		(-3.69)	-INF	-3.23		75% 11%	0 Unscorable
2	-1.90	.04	-2.78	-3.23	-2.47	-2.78	34% 41%	2 Minimal
3	-2.21	.03	-2.20	-2.47	-1.91	-2.42	23% 28%	3
4	-3.07	.02	-1.54	-1.91	-1.04	-2.15	46% 26%	4 Limited
5	-.73	.01	-.42	-1.04	.34	-.96	35% 34%	5
6	-.21	.01	1.45	.34	2.62	.11	65% 74%	6 Adequate
7	3.07	.01	3.49	2.62	4.43	2.82	40% 48%	7
8	3.85	.01	(5.20)	4.43	+INF	4.15	71% 23%	8 Superior

Table 25: Grade 11 Threshold Difficulties – Prompt 3 Domain 3

Domain ITEM DIFFICULTY MEASURE OF -.04 ADDED TO MEASURES

CATEGORY LABEL	OBSERVED SCORE	OBSVD COUNT	SAMPLE %	INFINIT	OUTFIT	STRUCTURE	CATEGORY			
				MNSQ	MNSQ	MEASURE	MEASURE			
0	0	1001	1	-2.11	-2.41	1.51	1.79	NONE	(-3.83)	0 Unscorable
2	1	1185	1	-2.11	-1.88	.70	.70	-2.27	-2.74	2 Minimal
3	2	1519	1	-1.50	-1.21	.65	.61	-1.76	-2.07	3
4	3	14478	11	-.31	-.33	.95	.93	-3.00	-1.24	4 Limited
5	4	15462	12	.61	.68	.90	.88	.14	-.03	5
6	5	61950	48	1.64	1.68	.94	.91	-.17	1.61	6 Adequate
7	6	20551	16	2.81	2.70	.85	.82	3.33	3.51	7
8	7	12798	10	3.82	3.74	.92	.88	3.73	(5.11)	8 Superior
MISSING	1381	1	.15							

AVERAGE MEASURE is mean of measures in category.

+									
CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		
LABEL	MEASURE	S.E.	AT CAT.	----ZONE----		PROBABILITY	M->C	C->M	
-----+-----									
0	NONE		(-3.83)	-INF	-3.27		85%	8%	0 Unscorable
2	-2.31	.04	-2.74	-3.27	-2.39	-2.87	46%	40%	2 Minimal
3	-1.80	.03	-2.07	-2.39	-1.71	-2.30	25%	34%	3
4	-3.04	.02	-1.24	-1.71	-.63	-2.01	58%	32%	4 Limited
5	.11	.01	-.03	-.63	.63	-.41	29%	39%	5
6	-.21	.01	1.61	.63	2.70	.34	69%	72%	6 Adequate
7	3.29	.01	3.51	2.70	4.39	2.94	38%	52%	7
8	3.69	.01	(5.11)	4.39	+INF	4.07	77%	24%	8 Superior

Table 26: Grade 11 Threshold Difficulties – Prompt 3 Domain 3

Domain ITEM DIFFICULTY MEASURE OF .11 ADDED TO MEASURES

CATEGORY		OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY	
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE
0	0	1001	1	-2.11	-2.41	1.52	1.77	NONE	(-3.88)
2	1	1332	1	-2.11	-1.87	.66	.66	-2.52	-2.71
3	2	1858	1	-1.45	-1.22	.65	.60	-1.99	-2.01
4	3	11841	9	-.42	-.36	.86	.80	-2.76	-1.24
5	4	17995	14	.63	.68	.88	.87	-.38	-.12
6	5	69386	53	1.74	1.78	.91	.89	-.24	1.82
7	6	18592	14	3.11	2.91	.79	.79	3.55	4.07
8	7	6939	5	4.09	4.03	.94	.89	4.34	(5.80)
MISSING		1381	1	.14					

AVERAGE MEASURE is mean of measures in category.

+										
CATEGORY	STRUCTURE		SCORE-TO-MEASURE			50% CUM.	COHERENCE		OBS-EXP	
LABEL	MEASURE	S.E.	AT CAT.	----	ZONE----	PROBABLY	M->C	C->M	RESIDUAL	
+										
0	NONE		(-3.88)	-INF	-3.29		85%	8%	-1.2%	0 Unscorable
2	-2.41	.04	-2.71	-3.29	-2.34	-2.91	48%	41%	-.8%	2 Minimal
3	-1.88	.03	-2.01	-2.34	-1.67	-2.25	27%	33%	-.5%	3
4	-2.65	.02	-1.24	-1.67	-.72	-1.89	53%	32%	-.2%	4 Limited
5	-.27	.01	-.12	-.72	.60	-.60	35%	41%	.0%	5
6	-.12	.01	1.82	.60	3.17	.31	72%	79%	.1%	6 Adequate
7	3.66	.01	4.07	3.17	5.03	3.40	43%	46%	.1%	7
8	4.46	.01	(5.80)	5.03	+INF	4.75	70%	13%	.1%	8 Superior

Table 27: Grade 11 Threshold Difficulties – Prompt 3 Domain 4

Domain ITEM DIFFICULTY MEASURE OF .02 ADDED TO MEASURES

CATEGORY LABEL	OBSERVED SCORE	OBSVD COUNT	SAMPLE %	INFINIT	OUTFIT	STRUCTURE MEASURE	CATEGORY MEASURE
----------------	----------------	-------------	----------	---------	--------	-------------------	------------------

LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	1001	1	-2.11	-2.43	1.55	1.83	NONE	(-3.85)	0 Unscorable
2	1	1217	1	-2.21*	-1.90	.53	.48	-2.35	-2.72	2 Minimal
3	2	1747	1	-1.53	-1.25	.58	.50	-1.95	-2.04	3
4	3	11289	9	-.52	-.39	.76	.67	-2.71	-1.29	4 Limited
5	4	18186	14	.50	.64	.77	.74	-.38	-.17	5
6	5	65821	51	1.71	1.71	.80	.80	-.12	1.70	6 Adequate
7	6	20204	16	2.99	2.80	.78	.76	3.42	3.79	7
8	7	9479	7	3.98	3.88	.91	.86	4.08	(5.47)	8 Superior
MISSING		1381	1	.15						

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE	SCORE-TO-MEASURE	50% CUM.	COHERENCE	OBS-EXP	
LABEL	MEASURE S.E.	AT CAT. ----ZONE----	PROBABLT	M->C C->M	RESIDUAL	
0	NONE	(-3.85) -INF -3.27		85% 8%	-1.7%	0 Unscorable
2	-2.33 .04	-2.72 -3.27 -2.36	-2.88	48% 45%	-1.2%	2 Minimal
3	-1.93 .03	-2.04 -2.36 -1.70	-2.28	28% 32%	-.8%	3
4	-2.70 .02	-1.29 -1.70 -.77	-1.93	56% 32%	-.4%	4 Limited
5	-.36 .01	-.17 -.77 .56	-.66	39% 47%	-.1%	5
6	-.11 .01	1.70 .56 2.93	.29	72% 77%	.1%	6 Adequate
7	3.44 .01	3.79 2.93 4.72	3.15	42% 50%	.2%	7
8	4.10 .01	(5.47) 4.72 +INF	4.42	76% 20%	.3%	8 Superior

Table 28: Grade 11 Threshold Difficulties – Prompt 3 Domain 5

Domain ITEM DIFFICULTY MEASURE OF .03 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFIT	OUTFIT	STRUCTURE	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	MEASURE	MEASURE	
0	0	1001	1	-2.11	-2.40	1.49	1.77	NONE	(-3.86)	0 Unscorable
2	1	1225	1	-2.17*	-1.87	.56	.51	-2.36	-2.74	2 Minimal
3	2	1907	1	-1.46	-1.18	.62	.58	-2.00	-2.04	3
4	3	13994	11	-.30	-.28	.89	.86	-2.77	-1.20	4 Limited
5	4	19959	15	.69	.73	.91	.90	-.16	.05	5
6	5	57021	44	1.70	1.72	.92	.91	.15	1.71	6 Adequate
7	6	23300	18	2.87	2.75	.83	.86	3.10	3.63	7
8	7	10537	8	3.82	3.82	.97	.96	4.04	(5.39)	8 Superior
MISSING		1381	1	.15						

AVERAGE MEASURE is mean of measures in category.

CATEGORY	STRUCTURE	SCORE-TO-MEASURE	50% CUM.	COHERENCE	
LABEL	MEASURE S.E.	AT CAT. ----ZONE----	PROBABLT	M->C C->M	
0	NONE	(-3.86) -INF -3.29		85% 8%	0 Unscorable
2	-2.33 .04	-2.74 -3.29 -2.37	-2.89	47% 43%	2 Minimal
3	-1.97 .03	-2.04 -2.37 -1.68	-2.30	28% 33%	3
4	-2.73 .02	-1.20 -1.68 -.59	-1.94	55% 30%	4 Limited
5	-.13 .01	.05 -.59 .75	-.45	36% 43%	5
6	.18 .01	1.71 .75 2.76	.53	64% 72%	6 Adequate
7	3.13 .01	3.63 2.76 4.60	2.92	43% 48%	7
8	4.07 .01	(5.39) 4.60 +INF	4.33	68% 21%	8 Superior