Appendix 3.B Evaluating the Use of Item-Pattern and Number-Correct to Scale Score Scoring for Reporting Subscores

Maryland High School Assessment

Evaluating the Use of Item-Pattern and Number-Correct to Scale Score Scoring for Reporting Subscores

February 20, 2004

Educational Testing Service

Appendix 3.B Evaluating the Use of Item-Pattern and Number-Correct to Scale Score Scoring for Reporting Subscores

For the January 2004 administration of the Maryland High School Assessments, subscore scale scores were created using number-correct (NC) score to scale score conversion tables. However, the MSDE and the National Psychometric Council are interested in possibly reporting subscores based on item-pattern (IP) scoring, as will be used for reporting total test scores. While subscores will not be reported at the individual student level, the subscores will be aggregated at the classroom level to provide teachers and administrators with additional information about student performance by each of the reporting categories. To help determine the feasibility of implementing item-pattern scoring at the subscore level, this study investigates the nature and extent of differences in subscores based on item-pattern scoring versus number-correct scoring.

The results included in this report were based on the Algebra A04 form, which was administered this January. The distributions of items by type for each subscore (which were called Expectations) in Algebra A04 are listed in Table 3.B.1 below.

Table 3.B.1. Distribution of Items by Type for each Subscore

Reporting Category	Item Type				Total Points per Category
	ECR (4 pts/ECR)	BCR (3 pts/BCR)	SPR (1 pt/SPR)	SR (1 pt/SR)	
Expectation 1.1 The student will analyze a wide variety of patterns and functional relationships using the language of mathematics and appropriate technology.	1	0	1	8	13
Expectation 1.2 The student will analyze a wide variety of patterns and functional relationships using the language of mathematics and appropriate technology.	1	0	3	10	17
Expectation 3.1 The student will collect, organize, analyze, and present data.	0	2	2	4	12
Expectation 3.2 The student will apply the basic concepts of statistics and probability to predict possible outcomes of real-world situations.	1	1	0	4	11
TOTALS	3	3	6	26	53

Because item responses were not yet available for the Algebra A04 form, item responses were simulated based on 5000 simulees with a mean scale score of 398.36, standard deviation 43.18, using the existing "pre-equated"[7] item parameters for this form from the item bank.

7 The items were administered in either 2002 or 2003 – these item parameters were on the operational scale.

Item-pattern scale scores based on these item response vectors were then estimated for each subscore and for the total test. NC scale scores based on NC to scale score conversion tables were also produced for each subscore and the total score (see Appendix 3.B.a). Thus, each item response vector yielded 10 scale scores: a NC scale score and an IP scale score, for each of the four Expectations and the total test.

Results

Individual Scores

The mean scale scores for both the IP and NC scale scores were lower than the mean true scale scores (see Table 3.B.2). Whereas the true score ranged from 254 to 557, both the NC and IP scale scores ranged from 240 to 625; this is due to the assignment of the lowest and highest obtainable score (LOSS; HOSS) for both the NC and IP estimated scores.[8]

Comparing the mean IP and NC scale scores, with the exception of Expectation 3.2, the NC means were very close to the IP means with less than a scale score difference. For Expectation 3.2, the NC scale score was higher by 11.02 scale score points; this result is examined in detail later in this section. The smallest difference between the mean scores was Expectation 1.1 with a difference of only 0.12 scale score points. All of the NC scale score means were slightly higher than the IP scale score means, except for the total scale core. See Appendix 3.B.b for the number, percent, mean, and standard deviation of NC and IP scale scores grouped at intervals of 10 true scale score points for each of the Expectations and the total scale core (i.e., a tabled true score of 405 includes results for all true scale scores from 400 to 409). The standard error associated with selected IP scale scores from each distribution of scores is listed in Appendix 3.B.c

Table 3.B.2. Summary Statistics

Scale Score	Total		Expectation 1.1		Expectation 1.2		Expectation 3.1		Expectation 3.2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
True	398.36	43.18	-	-	-	-	-	-	-	-
NC	396.77	49.48	396.30	63.47	395.63	59.16	398.36	67.18	391.93	74.70
IP	397.11	48.78	396.18	63.08	395.46	59.50	398.13	67.63	380.91	89.54

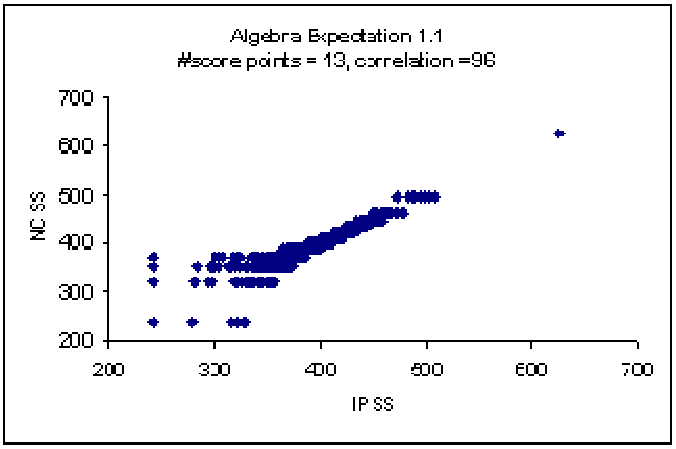
Not unexpectedly, the correlations between the IP and NC scale scores were high, ranging from .91 to .98 for the subscores and .98 for the total scale score. As noted by the bivariate plots (see Figures 3.B.1 –3.B.5) and the difference in standard deviation of NC and IP scale scores given the true scale score (see Figures 3.B.6-3.B.10), the largest differences in scores were noted at the lower end of the scale. This result is expected, given that the consideration

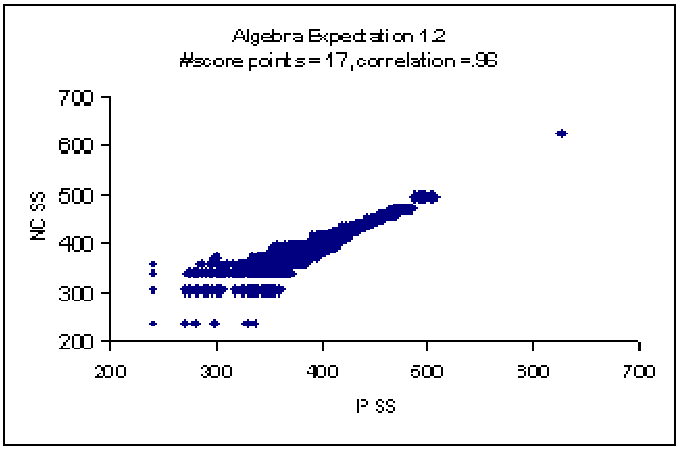
8 The LOSS and HOSS, which were assigned to extreme scores for which IRT does not provided maximum likelihood ability estimates, were set after examining the scale scores produced for the other scores.

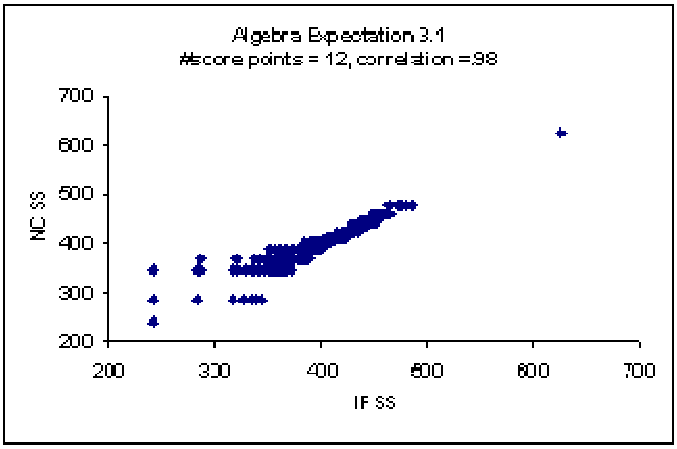
of the C-parameter ("guessing") has a greater effect among low-scoring examinees (Yen, 1984; Yen & Candell, 1991). The variation of scores was also greater at the lower end of scale for the total score, although the amount of variation was smaller than for the subscores. This result is also expected, given that as the number of score points increase, the influence of the uncertainty introduced by guessing decreases.

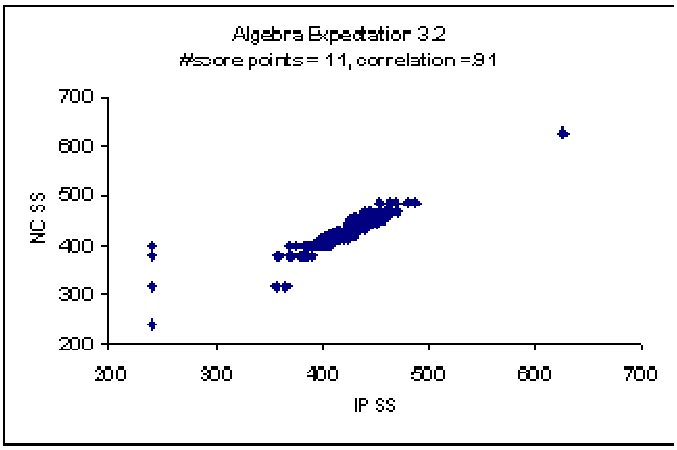
Appendix 3.B

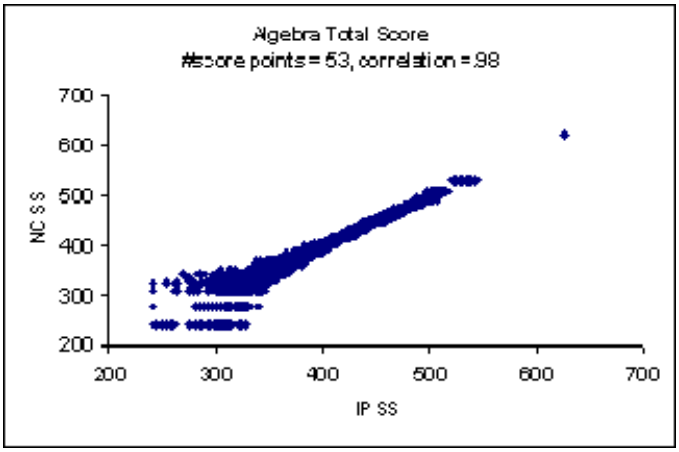
Figures 3.B.1 – 3.B.5 Bivariate Plots of NC and IP Scale Scores









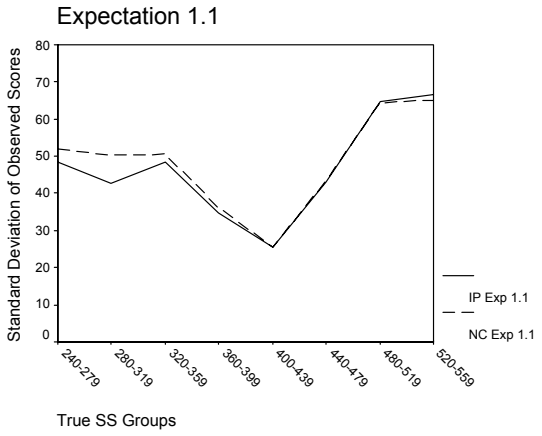


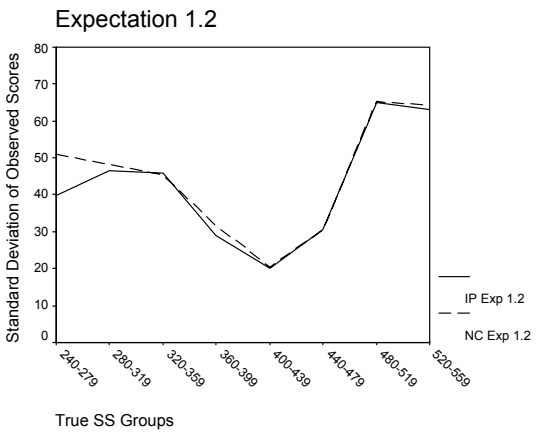
Following IRT principles, IP scale scores should have lower conditional standard errors of measurement than NC scale scores. This result is seen with the exception of Expectation 3.1

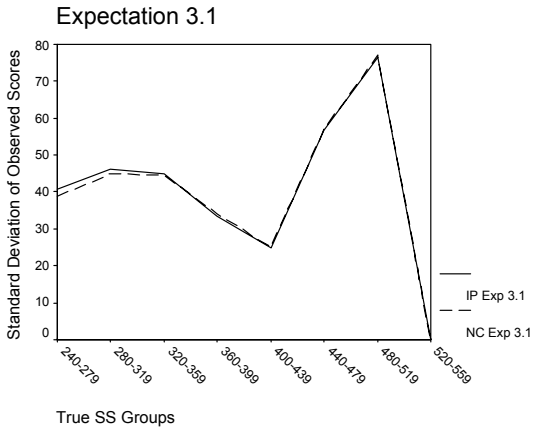
and 3.2 (see Figures 3.B.6-3.B.10). Both of these subscores have the fewest score points: Expectation 3.1 has 12 score points and Expectation 3.2 has only 11 score points. In both cases, the LOSS was assigned to more simulees using IP scoring compared to NC scoring (see Table 3.B.3). As noted in Table 2 of the two subscores, Expectation 3.2 has more variation and a larger difference in the average scale scores for the IP and NC scoring procedures. This is due to the large number of simulees that received the LOSS via IP scoring (n=1127) compared to the number of simulees that received the LOSS via NC scoring (384). In contrast, for Expectation 1.2, 119 simulees received the LOSS via IP scoring and 203 simulees received the LOSS via NC scoring. For this subscore, the NC scores were more variable than the IP scale scores and the difference in average scale scores was smaller.

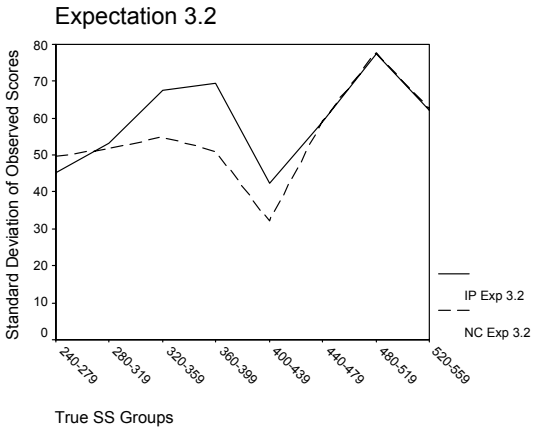
Appendix 3.B

Figures 3.B.6-3.B.10. Empirical Conditional Standard Errors of Scale Scores for Item Pattern (IP) and Number Correct (NC) Scoring Methods









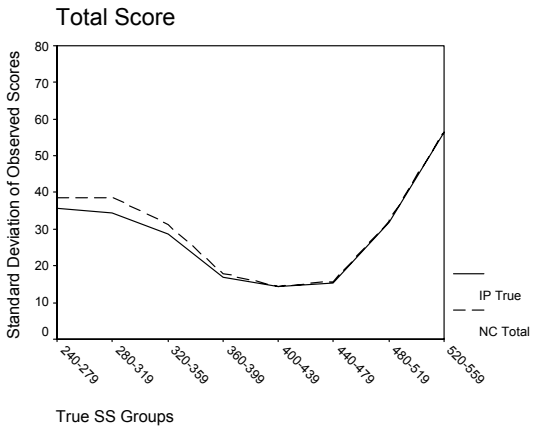


Table 3.B.3. Number and Percent of Simulees Assigned the LOSS by Subscore

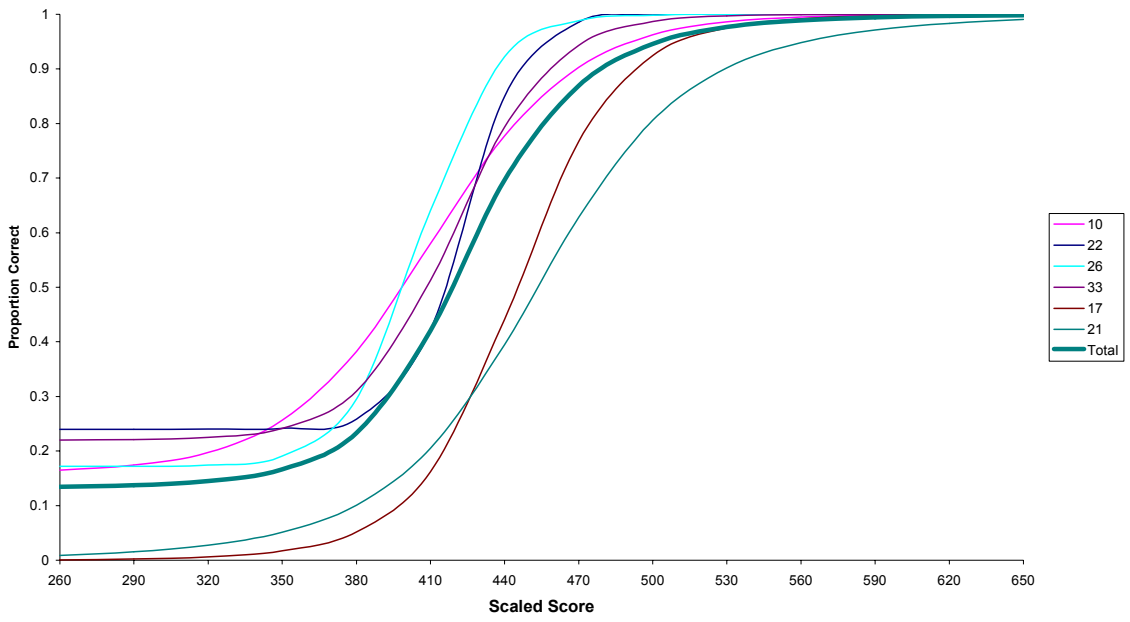
	IP		NC	
	N	%	N	%
Expectation 1.1	217	4.3	279	5.6
Expectation 1.2	119	4.0	203	4.1
Expectation 3.1	187	3.7	114	2.3
Expectation 3.2	1127	22.5	384	7.7
Total Score	66	1.3	80	1.6

Examining the IP and NC scale scores for Expectation 3.2 in more detail, it is noted that the differences in scores is related to characteristics of the items. This subscore included only 6 items: four selected response items (1 point each), one brief constructed response item (3 points) and one extended constructed response item (4 points). The SR items were moderately difficult, with B-values ranging between 406 and 424 (see Table 3.B.4) and have c values ranging from .16 to .24. In contrast, the BCR and ECR items were relatively more difficult, have 0 guessing, and contribute the most information (see Figure 3.B.11).

Table 3.B.4. Expectation 3.2 Item Parameters

		Parameters						
Item	Type	A	C	B	B-1	B-2	B-3	B-4
10		0.0212	0.16	410.04				
22	SR	0.0532	0.24	423.56				
26	SR	0.0393	0.17	406.02				
33	SR	0.0309	0.22	420.02				
17	BCR	0.0196			385.92	450.76	439.10	
21	ECR	0.0145			413.04	477.14	445.79	439.16

Figure 3.B.11. Expectation 3.2 Item Characteristic Curves and Expectation 3.2 Characteristic Curve



The effect of these item parameters on individual scores can be more clearly observed by examining the scores within the true score range of 320 to 359. In this score range, there were 33 possible IP scores compared to 6 possible NC scores (see Table 3.B.5).

Appendix 3.B

Table 3.B.5. Distribution of IP and NC Scale Scores for Expectation 3.2 within the True Score Grouping 320-359

Scale Score	IP		NC	
	N	%	N	N%
240	455	62.33	171	23.42
316			300	41.10
357	36	4.93		
358	11	1.51		
359	13	1.78		
366	43	5.89		
368	16	2.19		
370	10	1.37		
375	5	0.68		
377	6	0.82	180	24.66
379	14	1.92		
382	14	1.92		
383	9	1.23		
386	13	1.78		
387	9	1.23		
388	10	1.37		
390	6	0.82		
391	13	1.78		
394	5	0.68		
395	7	0.96		
398			65	8.90
400	3	0.41		
401	1	0.14		
402	3	0.41		
403	4	0.55		
404	1	0.14		
405	1	0.14		
406	7	0.96		
407	2	0.27		
408	2	0.27		
411			11	1.51
412	2	0.27		
414	2	0.27		
420	3	0.41		
421	2	0.27		
422			3	0.41
423	2	0.27		

Based on IP scoring, the LOSS (240) was assigned for all response patterns where only one, two, or three score points were obtained on the SR items. In contrast, one score point obtained on either the BCR or the ECR resulted in a much higher IP scale score: 366 and 357, respectively (see Table 3.B.6). This result is due to the item pattern scoring process: if a simulee gets 3 or less points from SR items, but 0 from the BCR or ECR items, the item pattern scoring process concludes that these points were likely to have come from guessing, and the IP scale score is at the LOSS. However, when a score point is obtained from a BCR or ECR item, the item pattern scoring process concludes that this score point was obtained via knowledge, not guessing, and the IP scale score is substantially higher than the LOSS.

Table 3.B.6. Expectation 3.2 Item Pattern Response Patterns and Associated IP and NC Scale Scores

			Items					
IP Scale Score	Raw Score	NC Scale Score	10 (SR)	22 (SR)	26 (SR)	33 (SR)	17 (BCR)	21 (ECR)
240	1	316	0	0	0	1	0	0
240	1	316	0	0	1	0	0	0
240	1	316	0	1	0	0	0	0
240	1	316	1	0	0	0	0	0
366	1	316	0	0	0	0	1	0
357	1	316	0	0	0	0	0	1
240	2	377	0	0	1	1	0	0
240	2	377	0	1	1	0	0	0
240	2	377	0	1	0	1	0	0
240	2	377	1	1	0	0	0	0
240	3	398	0	1	1	1	0	0

To shed further light on how IP scale scores were related to the NC scale scores for each subscore and the total score, the IP scale scores were grouped by the corresponding NC scale score and the following statistics were computed (see Tables 3.B.7 to 3.B.11):

1. Number of scores within the grouping (N)
2. Mean IP scale score (Mean)
3. Standard deviation IP scores (SD)
4. Number of IP scale scores within 5 Scale Scores of the NC scale score (N within 5 NC SS)
5. Percent of IP scale scores within 5 Scale Scores of the NC scale score (N within 5 NC SS)
6. Minimum obtained IP scale score (Low)
7. Maximum obtained IP scale score (High)
8. Mean IP scale score standard error (AveSE)

Table 3.B.7. Expectation 1.1

Raw Score	NC Scale Score	Pattern Scores							
		N	Mean	SD	N Within 5 NC SS	% Within 5 NC SS	Low	High	AveSE
0, 1	240	279	267.06	34.65	162	58.06%	240	328	95.57
2	320	388	314.67	34.31	52	13.40%	240	356	52.02
3	352	506	345.98	24.05	160	31.62%	240	372	34.40
4	372	535	367.99	14.42	256	47.85%	240	385	26.42
5	387	548	386.73	5.52	385	70.26%	364	399	21.65
6	400	503	399.62	4.59	375	74.55%	386	411	19.29
7	411	462	410.80	4.16	375	81.17%	401	421	18.00
8	422	405	422.18	4.28	332	81.98%	411	431	17.65
9	433	373	433.19	4.28	300	80.43%	422	444	18.36
10	446	380	445.74	5.16	246	64.74%	433	459	20.49
11	464	311	464.31	6.15	149	47.91%	449	478	25.66
12	494	228	491.48	8.48	165	72.37%	472	508	35.50
13	625	82	625.00	0.00	82	100.00%	625	625	206.86

Table 3.B.8. Expectation 1.2

Raw Score	NC Scale Score	Pattern Scores							
		N	Mean	SD	N Within 5 of NC SS	% Within 5 of NC SS	Low	High	AveSE
0,1	240	203	263.53	33.96	125	61.58%	240	336	162.30
2	306	284	307.40	39.46	7	2.46%	240	360	78.49
3	343	385	336.90	30.14	79	20.52%	240	369	40.86
4	361	448	359.20	17.24	148	33.04%	240	384	24.33
5	375	482	373.05	10.95	212	43.98%	300	394	19.68
6	386	419	385.46	8.88	200	47.73%	351	404	17.80
7	397	427	395.12	7.91	229	53.63%	357	412	16.92
8	406	395	405.99	6.11	226	57.22%	388	418	16.33
9	416	381	414.79	5.63	266	69.82%	392	424	16.09
10	425	337	424.34	4.65	276	81.90%	409	434	15.93
11	434	287	433.56	3.78	250	87.11%	419	441	15.81
12	443	234	442.10	2.98	216	92.31%	432	449	15.85
13	452	197	451.81	2.59	188	95.43%	442	458	16.39
14	462	181	462.15	2.44	177	97.79%	455	469	17.78
15	475	149	475.61	2.66	144	96.64%	467	485	20.80
16	496	127	496.48	4.14	107	84.25%	487	505	27.89

17	625	64	625.00	0.00	64	100.00%	625	625	286.66

Appendix 3.B

Table 3.B.9. Expectation 3.1

Raw Score	NC Scale Score	Pattern Scores							
		N	Mean	SD	N Within 5 of NC SS	% Within 5 of NC SS	Low	High	AveSE
0	240	114	240.00	0.00	114	100.00%	240	240	205.53
1	283	344	297.65	32.30	83	24.13%	240	344	83.36
2	344	606	338.94	24.92	176	29.04%	240	370	39.29
3	369	656	366.77	14.56	226	34.45%	286	387	26.19
4	387	627	385.33	8.86	327	52.15%	351	400	22.00
5	402	519	400.86	5.09	371	71.48%	382	410	19.39
6	413	462	413.66	3.30	418	90.48%	405	423	17.59
7	424	363	423.98	3.73	312	85.95%	416	433	16.64
8	434	362	434.06	3.92	290	80.11%	426	442	16.39
9	445	301	445.20	4.30	228	75.75%	437	453	17.17
10	458	248	457.46	4.92	171	68.95%	447	465	19.47
11	479	225	477.84	5.82	191	84.89%	463	484	26.80
12	625	173	625.00	0.00	173	100.00%	625	625	379.47

Table 3.B.10. Expectation 3.2

Raw Score	NC Scale Score	Pattern Scores							
		N	Mean	SD	N Within 5 of NC SS	% Within 5 of NC SS	Low	High	AveSE
0	240	384	240.00	0.00	384	100.00%	240	240	242.76
1	316	823	279.51	57.42	0	0.00%	240	366	178.03
2	377	832	350.16	55.80	212	25.48%	240	391	74.31
3	398	702	391.69	26.83	361	51.42%	240	407	28.92
4	411	530	412.86	6.27	290	54.72%	398	423	17.42
5	422	430	424.13	6.03	226	52.56%	406	432	15.84
6	432	345	433.90	6.30	106	30.72%	414	441	15.91
7	442	248	443.01	6.67	109	43.95%	424	450	16.83
8	453	197	452.52	7.36	80	40.61%	429	459	18.44
9	466	175	462.22	7.34	134	76.57%	440	470	20.80
10	484	164	481.70	7.76	148	90.24%	454	488	28.90
11	625	170	625.00	0.00	170	100.00%	625	625	328.74

Appendix 3.B

Table 3.B.11. Total Test

Raw Score	NC Scale Score	Pattern Scores							
		N	Mean	SD	N Within 5% NC SS	% Within 5 NC SS	Low	High	AveSE
0-4	240	80	271.51	30.78	35	43.75%	240	326	59.62
5	281	70	298.97	29.61	2	2.86%	240	339	38.29
6	309	99	303.64	35.00	14	14.14%	240	345	37.58
7	324	136	324.01	19.18	36	26.47%	240	349	23.55
8	335	135	334.87	15.60	36	26.67%	273	356	19.63
9	344	145	340.93	15.55	56	38.62%	269	363	18.08
10	351	159	350.57	10.80	65	40.88%	304	368	15.64
11	358	164	356.62	7.40	98	59.76%	336	370	14.44
12	363	184	362.86	6.36	123	66.85%	337	380	13.46
13	368	179	366.70	6.87	121	67.60%	337	380	12.94
14	373	183	372.50	5.50	139	75.96%	338	383	12.17
15	377	187	377.14	4.86	143	76.47%	362	388	11.62
16	382	143	380.78	4.47	115	80.42%	364	390	11.21
17	385	158	384.96	4.15	134	84.81%	372	393	10.77
18	389	158	388.32	3.32	139	87.97%	380	397	10.43
19	392	151	392.21	3.60	130	86.09%	381	400	10.07
20	396	129	395.45	3.32	120	93.02%	381	403	9.79
21	399	142	398.15	3.04	132	92.96%	389	405	9.56
22	402	125	401.66	2.87	118	94.40%	394	409	9.29
23	405	111	404.69	3.03	105	94.59%	396	412	9.08
24	408	112	408.00	2.41	112	100.00%	403	413	8.86
25	410	122	410.50	2.29	121	99.18%	405	416	8.71
26	413	111	413.10	2.30	111	100.00%	408	418	8.57
27	416	111	415.43	2.43	107	96.40%	410	422	8.46
28	418	107	418.37	2.71	103	96.26%	412	425	8.34
29	421	120	421.12	2.51	118	98.33%	414	426	8.26
30	423	108	423.51	2.45	106	98.15%	418	430	8.21
31	426	84	425.73	2.72	82	97.62%	420	432	8.18
32	428	97	428.47	2.56	93	95.88%	421	434	8.17
33	431	84	430.68	2.67	79	94.05%	424	435	8.18
34	433	83	433.66	2.30	82	98.80%	428	439	8.23
35	436	80	436.29	2.35	78	97.50%	429	441	8.29
36	439	84	438.63	2.25	83	98.81%	432	444	8.37
37	441	78	441.13	2.54	77	98.72%	435	446	8.48
38	444	65	444.03	2.49	64	98.46%	437	448	8.62

Raw Score	NC Scale Score	Pattern Scores							
		N	Mean	SD	N Within 5% NC SS	% Within 5 NC SS	Low	High	AveSE
39	447	64	447.19	2.56	61	95.31%	438	453	8.82
40	449	72	449.29	2.71	70	97.22%	440	454	8.97
41	452	63	452.29	2.88	60	95.24%	444	458	9.22
42	456	63	454.92	2.61	60	95.24%	449	461	9.46
43	459	69	459.29	2.43	67	97.10%	453	465	9.92
44	462	62	462.23	2.81	61	98.39%	456	467	10.28
45	466	48	466.31	2.49	47	97.92%	461	472	10.84
46	471	39	470.31	2.18	37	94.87%	465	475	11.45
47	476	44	474.84	2.57	42	95.45%	467	481	12.23
48	481	45	481.09	3.32	40	88.89%	475	488	13.46
49	488	45	488.20	2.64	45	100.00%	483	493	15.03
50	497	35	497.69	3.73	30	85.71%	489	507	17.51
51	510	37	508.41	4.18	29	78.38%	498	516	20.86
52	532	18	530.61	6.33	12	66.67%	521	544	30.23
53	625	12	625.00	0.00	12	100.00%	625	625	139.03

Note that regression effects affect these results: because simulees were grouped on the basis of an observed score (NC scale score), the dependent observed score (IP scale score) tends to be less extreme. Near the top and bottom of the scale, the means and standard deviations were also affected by the LOSS and HOSS.

Based on these tables, the mean IP scale score was similar to the NC scale score for the majority of the score groupings. As was observed in the true score groupings, the largest differences were noted at the lower end of the scale where the most variation of IP scale scores is also observed. In addition, the majority of the IP scale scores were within 5 scale score points of the NC scale score.

Aggregate Scores

As the primary purpose of the reported subscores will be to provide reports at the classroom level, aggregate scores were also simulated. To create these simulated results, 100 "classrooms" were simulated by randomly selecting 30 scores for each "classroom". These results are summarized in Table 3.B.12. The pattern of results is similar to the scores aggregated across the total sample (see Table 3.B.2). As with the total sample, the differences between the two types of scores were relatively small (less than one score point), with the exception of Expectation 3.2, where the NC scale scores were, on average, 10.24 points higher than the mean IP scale scores (see Table 3.B.13). The differences in IP and NC scale scores for each subscore are also observable in the bivariate plot (Figure 3.B.12).

Appendix 3.B

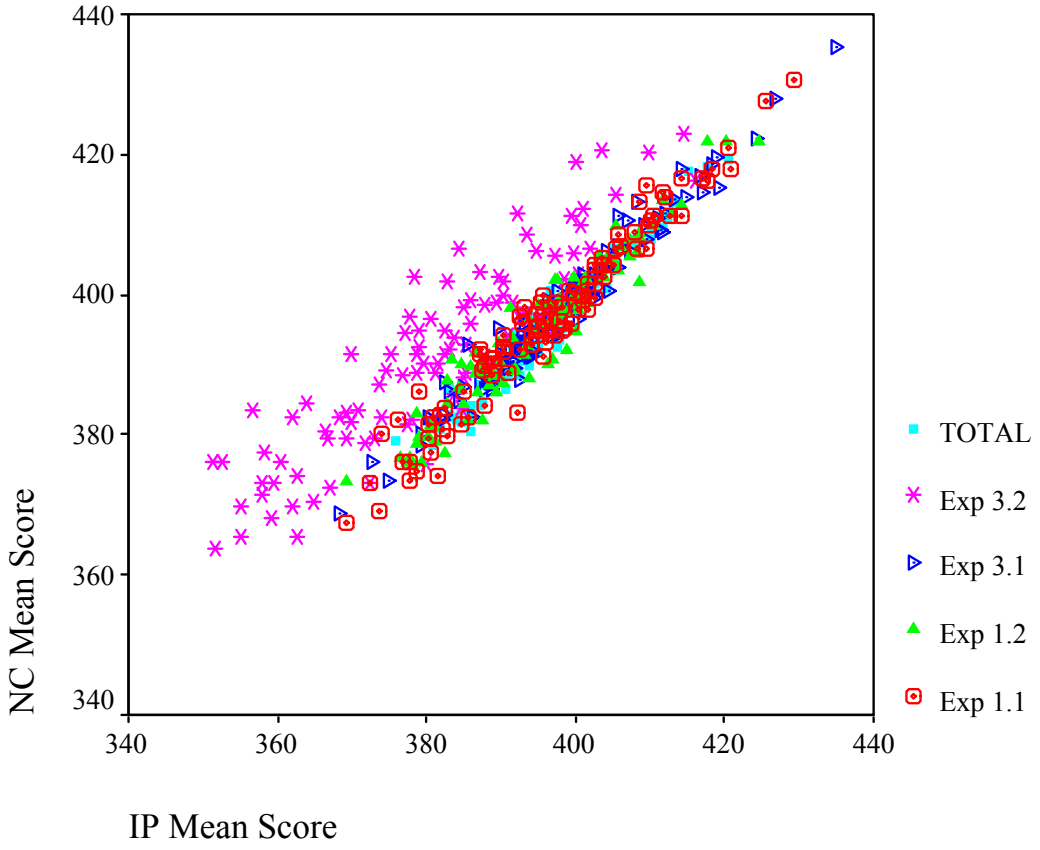
Table 3.B.12. Simulation of Aggregate Scores (n=30, 100 replications)

		Score Points	Mean	SD	Minimum	Maximum
Expectation 1.1	NC	13	396.23	13.01	367.57	430.47
	IP		396.16	12.61	369.40	429.27
Expectation 1.2	NC	17	394.67	10.77	373.40	422.10
	IP		394.63	10.85	369.10	424.70
Expectation 3.1	NC	12	397.96	11.64	368.83	435.17
	IP		398.00	11.73	368.27	435.03
Expectation 3.2	NC	11	391.37	13.54	363.70	422.87
	IP		381.13	14.94	351.47	415.97
Total Score	NC	53	396.04	9.26	378.60	419.37
	IP		396.78	8.92	375.90	420.57

Table 3.B.13. Differences between Mean IP and NC Scores (IP – NC)

	Mean	SD	Minimum	Maximum
Expectation 1.1	-0.07	2.88	-7.30	9.00
Expectation 1.2	-0.03	2.92	-7.54	6.64
Expectation 3.1	0.05	2.22	-7.20	4.60
Expectation 3.2	-10.24	6.53	-26.80	4.07
TOTAL	0.73	1.77	-3.94	5.30

Figure 3.B.12. Bivariate Plots IP and NC Mean Scores (n=30, 100 replications)



Summary and Conclusions

Based on the results of this study, the mean IP scale score was similar to the NC scale score for the total sample of the total score and all of the subscores except Expectation 3.2. For Expectation 3.2 the mean NC scale score was 11.02 scale score points higher than the mean IP scale score. For the samples of 30 scores, the mean IP and NC scores were similar across 100 replications except Expectation 3.2. In this case, the NC scale score was 10.24 points higher than the IP scale score.

The point of doing IP scoring is to benefit from a reduced conditional standard error of measurement relative to NC scoring. However, for the subscore with the fewest score points, Expectation 3.2, IP scale scores had much higher conditional SEMs than NC scores through the lower part of the score scale. This occurred because a much larger number of scores were assigned the LOSS using IP scoring compared to NC scoring. The difference in results was caused by differential "interpretation" by the IP and NC scoring methods of low scores that did/did not include score points earned on constructed response items. This study cannot determine the relative validity or meaningfulness of the scores produced by the IP and NC scoring methods, but only note that they can produce very different results when there are a small number of score points that include both SR and CR items.

It can also be noted that at the classroom level, which is where these scores are to be used, the IP and NC scoring methods produced nearly identical means—except for Expectation 3.2. Consistent IP and NC results at the group level reflect their tau-equivalence, which has been found in many other tests (Yen, 1984; Yen & Candell, 1991). In essence, the theoretical improvement in conditional SEM can be very useful for individual examinees, but is of no apparent value for groups of 30 or more students. The possibility exists that for small numbers of items with a mixed format, IP scoring will produce higher conditional SEMs and very different mean scores than NC scoring. Thus, IP scoring does not appear uniformly beneficial for subscores with small numbers of items with mixed formats.

This study demonstrates that conclusions about "areas of need" can be affected by the type of scoring used when there are small numbers of items with mixed formats contributing to a subscore. While Total scale scores are quite stable across IP and NC scoring, Expectation scores based on small numbers of items can be significantly affected by scoring procedure. For example, based on Table 3.B.2 results, the conclusion would be drawn that Expectation 3.2 is a serious area of need when IP scoring is used, but only a modest area of need when NC scoring is used. If IP scoring is used for subscores, then additional explanatory information will be needed so that scores are interpreted appropriately.

References

Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. *Journal of Educational Measurement, 21*, 93-111.

Yen, W. M., & Candell, G. L. (1991). Increasing score reliability with item-pattern scoring: An empirical study in five score metrics. *Applied Measurement in Education, 4,* 209-228.

Appendix 3.B

Appendix 3.B.a

Number-Correct to Scale Score Scoring Tables

Expectation 1.1		
NC	Scale Score	SEM
0	240	80
1	240	80
2	320	42
3	352	30
4	372	25
5	387	22
6	400	19
7	411	18
8	422	18
9	433	18
10	446	20
11	464	25
12	494	36
13	625	80

Expectation 1.2		
NC	Scale Score	SEM
0	240	80
1	240	80
2	306	55
3	343	28
4	361	22
5	375	19
6	386	18
7	397	17
8	406	16
9	416	16
10	425	16
11	434	16
12	443	16
13	452	16
14	462	18
15	475	21
16	496	28
17	625	80

Expectation 3.1		
NC	Scale Score	SEM
0	240	80
1	283	80
2	344	32
3	369	25
4	387	22
5	402	19
6	413	18
7	424	17
8	434	16
9	445	17
10	458	19
11	479	27
12	625	80

Expectation 3.2		
NC	Scale Score	SEM
0	240	80
1	316	80
2	377	32
3	398	22
4	411	18
5	422	16
6	432	16
7	442	17
8	453	18
9	466	22
10	484	30
11	625	80

Appendix 3.B

Total Score		
NC	Scale Score	SEM
0	240	80
1	240	80
2	240	80
3	240	80
4	240	80
5	281	45
6	309	28
7	324	22
8	335	19
9	344	17
10	351	15
11	358	14
12	363	13
13	368	13
14	373	12
15	377	12
16	382	11
17	385	11
18	389	10
19	392	10
20	396	10
21	399	9
22	402	9
23	405	9
24	408	9
25	410	9
26	413	9
27	416	8
28	418	8
29	421	8
30	423	8
31	426	8
32	428	8
33	431	8
34	433	8

Total Score		
NC	Scale Score	SEM
36	439	8
37	441	8
38	444	9
39	447	9
40	449	9
41	452	9
42	456	10
43	459	10
44	462	10
45	466	11
46	471	12
47	476	12
48	481	13
49	488	15
50	497	17
51	510	21
52	532	31
53	625	80

Appendix 3.B.b
Grouped Frequency Distribution

The following tables list the number, percent, mean and standard deviation of NC and IP scores grouped at intervals of 10 true scale score points.

Expectation 1.1

True Scale Score (midpoint)	N	%	NC		IP	
			Mean	SD	Mean	SD
255	4	0.08	288.00	56.94	240.00	0.00
265	3	0.06	277.33	64.66	284.33	47.82
275	10	0.20	307.60	49.90	303.00	48.95
285	13	0.26	273.23	44.52	284.69	45.66
295	34	0.68	288.12	44.79	276.00	40.95
305	45	0.90	301.96	54.29	293.04	44.75
315	61	1.22	282.80	50.00	283.85	41.24
325	100	1.98	303.70	51.21	299.25	47.66
335	135	2.70	321.84	51.14	319.74	49.74
345	203	4.06	330.13	49.40	329.96	47.87
355	292	5.84	342.38	46.88	343.79	42.43
365	363	7.26	352.78	41.97	354.56	36.37
375	424	8.48	368.89	35.87	367.92	36.90
385	453	9.06	378.63	30.04	379.21	29.52
395	429	8.58	391.90	25.26	392.11	24.95
405	442	8.84	403.05	21.60	402.70	22.12
415	401	8.02	415.05	21.73	415.20	21.71
425	401	8.02	426.97	21.85	427.19	21.75
435	336	6.72	436.74	24.37	436.45	23.48
445	265	5.30	450.86	29.87	451.12	29.34
455	194	3.88	464.02	43.37	464.29	42.74
465	143	2.86	475.57	42.64	475.57	42.69
475	86	1.72	491.14	60.32	490.93	60.02
485	74	1.48	501.72	54.24	502.42	54.32
495	43	0.86	522.26	70.53	521.23	71.24
505	19	0.38	529.05	68.00	530.63	66.98
515	13	0.26	562.23	71.01	561.85	71.68
525	5	0.10	546.40	71.75	545.40	72.78
535	4	0.08	592.25	65.50	591.00	68.00
545	4	0.08	592.25	65.50	591.00	68.00
555	1	0.02	625.00	0.0	625.00	0.0

Appendix 3.B

Expectation 1.2

True Scale Score (midpoint)	N	%	NC		IP	
			Mean	SD	Mean	SD
255	4	0.08	282.25	51.07	254.75	29.50
265	3	0.06	262.00	38.11	254.00	24.25
275	10	0.20	299.80	55.26	290.30	42.42
285	13	0.26	302.69	47.82	289.54	46.76
295	34	0.68	291.21	49.04	287.00	46.38
305	45	0.90	290.00	46.70	281.56	44.72
315	61	1.22	304.66	48.65	300.07	47.15
325	100	1.98	305.82	46.47	291.78	45.08
335	135	2.70	324.55	46.33	319.58	44.31
345	203	4.06	327.33	47.25	328.53	46.10
355	292	5.84	344.12	38.31	343.72	38.68
365	363	7.26	354.39	36.09	354.70	33.45
375	424	8.48	366.85	31.22	368.57	28.03
385	453	9.06	380.82	24.17	381.71	19.72
395	429	8.58	391.70	21.86	392.56	20.13
405	442	8.84	403.80	18.00	403.66	17.05
415	401	8.02	414.43	17.32	415.01	16.37
425	401	8.02	423.28	17.03	423.66	16.36
435	336	6.72	432.96	17.67	433.46	17.14
445	265	5.30	445.16	17.31	445.28	17.37
455	194	3.88	455.75	16.92	455.57	16.73
465	143	2.86	466.07	25.77	466.62	25.85
475	86	1.72	496.83	51.56	497.57	51.33
485	74	1.48	498.57	49.93	499.07	49.79
495	43	0.86	529.42	68.22	530.26	67.83
505	19	0.38	576.37	65.58	576.37	65.56
515	13	0.26	565.46	66.93	565.38	67.03
525	5	0.10	547.60	70.66	547.60	70.70
535	4	0.08	592.75	64.50	595.00	60.00
545	4	0.08	592.75	64.50	593.00	64.00
555	1	0.02	625.00	0.0	625.00	0.0

Appendix 3.B

Expectation 3.1

True Scale Score (midpoint)	N	%	NC		IP	
			Mean	SD	Mean	SD
255	4	0.08	313.50	35.22	298.50	39.03
265	3	0.06	303.33	35.22	297.33	50.46
275	10	0.20	307.40	44.57	291.60	43.45
285	13	0.26	282.46	44.80	261.31	37.10
295	34	0.68	296.09	42.31	285.79	44.02
305	45	0.90	292.58	49.26	286.84	46.67
315	61	1.22	299.30	43.46	300.25	46.23
325	100	1.98	308.20	47.05	305.27	46.37
335	135	2.70	323.70	42.63	321.33	43.27
345	203	4.06	328.80	45.34	325.30	47.31
355	292	5.84	344.58	39.55	344.77	37.57
365	363	7.26	354.13	36.72	355.59	34.24
375	424	8.48	365.77	35.00	366.67	34.53
385	453	9.06	378.43	28.64	378.16	29.01
395	429	8.58	391.95	24.04	392.24	23.99
405	442	8.84	402.10	22.08	402.27	22.22
415	401	8.02	413.48	20.23	413.51	20.22
425	401	8.02	424.22	18.96	424.09	18.68
435	336	6.72	435.86	25.84	436.13	25.38
445	265	5.30	448.58	31.98	448.61	31.73
455	194	3.88	469.22	48.32	469.24	48.09
465	143	2.86	492.04	65.87	492.40	65.56
475	86	1.72	520.55	75.97	520.67	75.85
485	74	1.48	545.07	78.85	545.88	77.97
495	43	0.86	555.63	75.45	555.47	75.69
505	19	0.38	569.00	75.56	569.84	74.65
515	13	0.26	567.23	76.25	567.92	75.33
525	5	0.10	625.00	0.00	625.00	0.00
535	4	0.08	625.00	0.00	625.00	0.00
545	4	0.08	625.00	0.00	625.00	0.00
555	1	0.02	625.00	0.0	625.00	0.0

Appendix 3.B

Expectation 3.2

True Scale Score	N	%	NC		IP	
			Mean	SD	Mean	SD
255	4	0.08	297.00	38.00	240.00	0.00
265	3	0.06	240.00	0.00	240.00	0.00
275	10	0.20	315.10	49.43	267.30	57.63
285	13	0.26	290.31	52.85	240.00	0.00
295	34	0.68	308.00	50.86	277.50	59.21
305	45	0.90	307.60	49.80	274.80	58.60
315	61	1.22	297.54	53.89	263.11	49.90
325	100	1.98	314.91	55.07	270.98	57.34
335	135	2.70	315.44	55.85	287.51	67.09
345	203	4.06	324.67	55.64	294.62	68.03
355	292	5.84	326.62	53.59	299.11	69.35
365	363	7.26	341.65	56.07	324.44	71.45
375	424	8.48	349.30	53.05	333.68	70.91
385	453	9.06	367.30	45.28	353.99	65.99
395	429	8.58	382.40	39.78	371.10	60.09
405	442	8.84	395.64	36.13	391.81	48.99
415	401	8.02	410.08	24.85	405.78	41.47
425	401	8.02	422.60	19.12	421.61	26.02
435	336	6.72	436.91	29.59	437.71	32.31
445	265	5.30	453.42	42.18	453.65	42.84
455	194	3.88	470.89	55.81	471.80	54.75
465	143	2.86	485.34	65.01	486.52	64.14
475	86	1.72	515.35	74.15	515.59	73.85
485	74	1.48	534.70	78.19	534.82	77.90
495	43	0.86	539.49	78.13	539.70	77.76
505	19	0.38	564.68	72.77	564.00	73.61
515	13	0.26	601.92	56.45	602.54	54.95
525	5	0.10	596.80	63.06	597.60	61.27
535	4	0.08	550.00	86.91	550.00	86.70
545	4	0.08	625.00	0.00	625.00	0.00
555	1	0.02	625.00	0.0	625.00	0.0

Appendix 3.B

Total Score

True Scale Score	N	%	NC		IP	
			Mean	SD	Mean	SD
255	4	0.08	284.75	32.62	240.00	0.00
265	3	0.06	240.00	0.00	275.67	31.56
275	10	0.20	314.20	28.46	301.30	28.85
285	13	0.26	279.15	41.19	284.15	33.75
295	34	0.68	288.53	36.84	290.71	34.44
305	45	0.90	296.00	39.66	291.69	34.07
315	61	1.22	294.79	38.51	299.48	35.12
325	100	1.98	310.48	36.91	306.20	35.13
335	135	2.70	330.35	32.49	329.18	30.91
345	203	4.06	337.35	28.33	340.85	19.98
355	292	5.84	349.97	22.80	352.24	17.80
365	363	7.26	361.12	16.52	361.97	14.00
375	424	8.48	372.21	13.90	372.70	13.29
385	453	9.06	382.72	12.31	382.85	11.42
395	429	8.58	393.85	11.15	393.89	10.82
405	442	8.84	404.14	10.15	404.12	9.73
415	401	8.02	414.60	9.36	414.87	8.94
425	401	8.02	424.19	8.65	424.26	8.50
435	336	6.72	433.84	9.34	434.13	8.94
445	265	5.30	445.14	9.93	445.17	9.40
455	194	3.88	455.11	10.35	455.20	9.77
465	143	2.86	463.84	12.28	464.12	11.44
475	86	1.72	478.17	15.32	478.28	15.21
485	74	1.48	486.54	17.28	486.82	16.72
495	43	0.86	495.63	25.99	495.28	25.50
505	19	0.38	517.16	33.26	517.53	33.04
515	13	0.26	542.69	58.19	542.46	58.16
525	5	0.10	537.40	49.89	537.20	50.66
535	4	0.08	546.50	54.87	548.00	52.62
545	4	0.08	596.25	57.50	595.75	58.50
555	1	0.02	625.00	0.0	625.00	0.0

Appendix 3.B.c

Pattern Scoring Standard Error of Measurement for Selected IP Scores

Expectation 1.1	
IP	IP SEM
240	126
280	69
290	59
300	54
320	42
330	38
340	34
350	31
360	28
370	26
380	23
390	21
400	19
410	18
420	18
430	18
440	19
450	21
460	24
470	27
480	31
489	34
493	36
499	38
503	40
625	207

Expectation 1.2	
IP	IP SEM
240	220
279	97
281	93
291	75
300	62
310	50
320	41
330	34
340	29
350	25
360	22
370	20
380	18
390	17
400	17
410	16
420	16
430	16
440	16
450	16
460	17
470	19
480	22
492	26
498	28
504	31
625	287

Expectation 3.1	
IP	IP SEM
240	206
283	87
286	82
320	44
340	33
351	29
360	27
370	25
380	23
390	21
400	20
410	18
420	17
430	16
440	17
450	18
460	20
471	24
484	30
625	379

Expectation 3.2	
IP	IP SEM
240	243
357	46
366	39
370	37
379	31
382	29
390	25
400	21
410	18
420	16
430	15
440	16
450	18
461	20
470	23
480	28
488	32
625	329

Total	
IP	IP SEM
240	90
251	75
260	64
269	55
273	51
280	45
290	38
300	32
310	27
320	23
330	20
340	18
350	15
360	14
370	12
380	11
390	10
400	9
410	9
420	8
430	8
440	8
450	9
460	10
470	11
480	13
490	15
500	18
510	21
521	26
530	30
540	35
625	139