

Variable Pass Rates Resulting from Equating Short Tests

John A. Stahl and James Masters

Pearson VUE

Paper Presented at the Annual Meeting of the American Educational Research
Association, San Diego, CA (2009)

Variable Pass Rates Resulting from Equating Short Tests

In certification and licensure testing, there are many situations where a program has tests with 30 or fewer items on a form. The domain of the area being tested may be limited in scope. The agency issuing the certification or license may specify that only a constrained set of topics will be covered by the examination. As more and more agencies enter the certification field, the number of these constrained testing conditions can be expected to increase.

Because the scope of a domain is limited, the number of questions that can be written to cover that domain is likewise limited. This limits the size of the item pool that can be used to construct a test form. The span of item difficulties represented in the item pool is restricted, and the distribution of difficulties across this span can be erratic. The necessity of controlling for enemy items and item exposure introduces further complications.

In these “small test, limited item pool” situations it can be very difficult to build two or more forms that are appropriately equated. According to Lord’s equity principle, “If an equating of tests x and y is to be equitable to each applicant, it must be a matter of indifference to applicants at every given level θ whether they are to take test x or test y .” However, in situations where a program has very short exams with a limited pool of items, the pass rate can be markedly affected by the choice of items. Two tests can meet the criteria of having the same cut score, based on IRT equating procedures and, in fact, have the same raw score cut score. However, the *effective cut score* can be significantly different on the two forms. Effective cut score is defined here as the latent ability measure that an individual has to meet the pass standard on a particular test form. It can

be shown that effective cut score is a function of the standard error at the cut point for the exam. Experience with many testing programs has shown that on short forms the effective cut score on one form can be different than the effective cut on another form constructed to the same test specifications. Clearly, in this situation it would not be a matter of indifference to the candidates as to which test they would prefer to take.

The goals of developing alternative forms of an examination include minimizing the degree of item overlap between forms, reducing the amount of item exposure, and minimizing the possibility of cheating. Different sets of items are chosen for each form under a variety of constraints. These constraints include meeting test blueprint guidelines, minimizing items on alternate forms, and meeting statistical targets. Statistical targets generally focus on achieving desired test information, difficulty means and distributions, and insuring that test equating requirements are met.

When the item pool is not robust, the alternate test forms can sometimes exhibit substantially different pass/fail rates, despite the best efforts of the test developer to meet the statistical targets.

The factors that can affect pass rate on parallel forms include:

- Effective logit cut
- Test length
- Distribution of item difficulties on the forms as compared to logit cut

In this paper, some of the factors that affect the pass rate on parallel forms and the ability of test developers to build parallel test forms are examined. Simulations are used to determine the effect of each of these factors. At the end of the paper, recommendations on how to minimize the adverse affects of these factors are offered.

Methods

The study used simulated data that was constructed to emulate live data. The items have IRT statistics (Rasch difficulty calibrations) assigned to them. Parallel forms were constructed using an automated test assembly (ATA) tool. The ATA tool constructs forms to match an ideal information function, while imposing the desired test specifications. The items were drawn from item pools of varying sizes and difficulty distributions. The size of the item pool compared the length of the test was also varied. The forms followed test blueprint constraints that are typical of a certification/licensure test. The forms were equated to a predetermined logit cut score but were not constrained to have identical raw cut scores.

Following the construction of the forms, a score-to-measure table was constructed for each of the forms. A score-to-measure table establishes the ability measure that corresponds to each individual raw score on the test, given the distribution of item difficulties for that test. An effective logit cut was determined from the score-to-measure table and was compared to the set logit cut that was used in the automated test assembly process. It is not uncommon for the effective logit cut to differ from the set logit cut and that this difference may vary between alternate forms, resulting in pass rate variation.

An example will clarify how this can come about. Assume that the set logit cut for a given testing program is 1.00 logits. Two 25-item forms are constructed using the ATA. Score-to-measure tables are also constructed for the two forms. On one form, a score of 18/25 has an effective logit value of 0.99 logits and a score of 19/25 has an effective logit value of 1.14 logits. The raw cut for this form would be set at 19 and the effective logit cut is 1.14 logits. On the second form, a score of 18/25 has a corresponding logit value of

.89 and a score of 19/25 has an effective logit value of 1.01. Again, the raw cut for the second form is set at 19; however, the effective logit cut on the second form is 1.01 logits. The difference in effective logit values for the two forms is .13 logits. This type of situation is not uncommon on short tests and can have a significant impact on the relative pass/fail rates across the two forms.

On longer tests, there is a smaller difference between effective logit values associated with a raw score of x and a raw score of $x-1$. The effect that such a difference can have was assessed by comparing the pass/fail rates determined by using the two effective logit values applied to a typical population distribution.

As a second part of the investigation, the impact that the distribution of items in the item pool has on the effective logit cuts was assessed. The distribution of item difficulties in the item pool was varied. Again, two forms were constructed using the same set of test specifications. A logit cut of 0.0 was used. The forms were not restricted to have the same raw score cut. Score-to-measure tables were constructed and the effective logit values were determined. The effective logit cut was then applied to a test population that consisted of 200 simulees, with their ability measures normally distributed and with a mean of 0.0 logits and a standard deviation (SD) of 1.0 logits,

Data

An ATA program assembled parallel forms from item banks with the following characteristics:

- Test length – 20, 30, 40, and 50 items
- Size of the item pool in relation to test length – 3, 5, and 7 times

- Distribution of item difficulties in the item pool (mean = 0, -1; SD = 0.5, 1.0, 1.5)
- Three content areas – A, B, and C. (Tests were built to have 25% A items, 35% B items, and 40% C items. The item pools were constructed to match test distribution.) This is the ideal situation. Departures from matching the test distribution impose further constraints on item selection. This issue was not explored in this paper.
- A logit cut score of 0.0 logits was used on all forms.

The data design in the Table 1 was used.

Table 1. Data Design

	Pool Characteristics		
Test Length	N	Mean	SD
X	3X	0	0.5
			1
			1.5
		-1	0.5
			1
			1.5
	5X	0	0.5
			1
			1.5
		-1	0.5
			1
			1.5
	7X	0	0.5
			1
			1.5
-1		0.5	
		1	
		1.5	

For each situation simulated, a score-to-measure table was determined for each form and the effective logit cuts identified.

The impact of differences in effective logit cuts was assessed against an examinee population in which the cut was at the mean of the examinee population. The population consisted of 200 individuals and had a mean of 0 and a SD of 1. This is the most extreme condition where the difference in effective cut scores would have the greatest impact on pass/fail rates.

Results

The tabulated results are presented in the Appendix. The tables are structured in the same manner as the data design. Two forms were created for each of the 18 test length/pool size combinations and the 3 conditions of item overlap. The raw score corresponding to a 0.0 logit cut is presented in the table, along with the effective logit cut and the number of simulees in the target population that would have passed at the resultant effective logit cut. The logit cut of 0.0 logits will always fall at a particular raw score or between two raw scores in any score-to-measure table. The effective logit cut will always, therefore, be equal to or greater than the logit cut. As can be seen in the tables in the Appendix, in most cases the effective logit cut is higher than the logit cut.

The focus of interest in this paper is on the impact of different effective cuts on pass rates. The statistic used is the difference in the number of individuals passing from the same simulee distribution.

The first variable is the length of the test. The average difference in the number of individuals by test length is presented in Table 2.

Table 2. Pass Rate by Test Length

Report

Difference in Number Who Pass

Number of Items on Test	Mean	N	Std. Deviation
20	5.94	54	5.537
30	3.39	54	3.779
40	2.59	54	2.962
50	2.65	54	2.549
Total	3.64	216	4.088

As the test length increases, the difference – across forms – in the number who pass decreases. This is a fairly obvious conclusion and the first reaction is to increase the length of the test to mitigate the impact. However, as was discussed in the introduction, the option to build a longer test is frequently not available.

The second variable is the size of the item pool relative to the test length. The results are presented in Table 3.

Table 3. Pass Rate by Pool Size Within Length of Test

Report

Difference in Number Who Pass

Number of Items on Test	Pool Size	Mean	N	Std. Deviation
20	60	8.39	18	6.409
	100	6.17	18	5.272
	140	3.28	18	3.578
	Total	5.94	54	5.537
30	90	4.78	18	4.519
	150	4.11	18	3.771
	210	1.28	18	1.602
	Total	3.39	54	3.779
40	120	2.72	18	2.697
	200	2.72	18	3.268
	280	2.33	18	3.049
	Total	2.59	54	2.962
50	150	3.44	18	3.148
	250	2.28	18	2.244
	350	2.22	18	2.074
	Total	2.65	54	2.549
Total	60	8.39	18	6.409
	90	4.78	18	4.519
	100	6.17	18	5.272
	120	2.72	18	2.697
	140	3.28	18	3.578
	150	3.78	36	3.440
	200	2.72	18	3.268
	210	1.28	18	1.602
	250	2.28	18	2.244
	280	2.33	18	3.049
	350	2.22	18	2.074
Total	3.64	216	4.088	

As expected, as the pool size increases, the difference in the number who pass decreases. Logically, the more items there are to choose from, the more likely it is that there will be smaller gaps in the item distribution between individual items on the test. The constraints that a limited domain place on item development may influence just how big a pool can be made. The degree to which items are excluded because of enemy relations will also decrease the effective pool size.

The third variable considers how well the item pool is targeted to the desired cut score. In this study, we used one pool that was ideally targeted to the cut score. The mean of the item distribution was set at 0.0 logits. The second pool was off-targeted by 1.0 logits. The mean of the item distribution was set at -1.0 logits. The results are presented in Table 4.

Table 4. Pass Rate by Pool Difficulty Within Test Length

Report

Difference in Number Who Pass

Number of Items on Test	Mean Logit of Pool	Mean	N	Std. Deviation
20	-1	5.41	27	6.535
	0	6.48	27	4.380
	Total	5.94	54	5.537
30	-1	2.44	27	1.908
	0	4.33	27	4.860
	Total	3.39	54	3.779
40	-1	2.56	27	2.792
	0	2.63	27	3.176
	Total	2.59	54	2.962
50	-1	3.00	27	2.760
	0	2.30	27	2.317
	Total	2.65	54	2.549
Total	-1	3.35	108	4.059
	0	3.94	108	4.116
	Total	3.64	216	4.088

Initially, these results were surprising. The off-targeted item pool generally had a smaller mean difference than the ideally targeted item pool. However, the off-targeted item pool had fewer items at the area of the cut score. Therefore, the forms constructed from this more limited set of items probably were very similar in difficulty distribution, and the logit gap between items was also likely similar. An off-targeted item pool does result in a larger error of measurement.

The next variable to be examined considered the spread of item difficulties within an item bank as measured by the SD of the item calibrations. The results are presented in Table 5.

Table 5. Pass Rate by Pool Standard Deviation by Pool Difficulty Within Test Length

Report

Difference in Number Who Pass

Number of Items on Test	Mean Logit of Pool	Std Dev of Pool	Mean	N	Std. Deviation
20	-1	.5	4.22	9	7.362
		1.0	7.44	9	8.079
		1.5	4.56	9	3.504
		Total	5.41	27	6.535
	0	.5	6.33	9	5.099
		1.0	6.67	9	4.472
		1.5	6.44	9	4.035
		Total	6.48	27	4.380
	Total	.5	5.28	18	6.238
		1.0	7.06	18	6.348
		1.5	5.50	18	3.792
		Total	5.94	54	5.537
30	-1	.5	3.11	9	2.088
		1.0	1.78	9	1.394
		1.5	2.44	9	2.128
		Total	2.44	27	1.908
	0	.5	4.11	9	5.302
		1.0	3.78	9	3.032
		1.5	5.11	9	6.194
		Total	4.33	27	4.860
	Total	.5	3.61	18	3.943
		1.0	2.78	18	2.510
		1.5	3.78	18	4.697
		Total	3.39	54	3.779
40	-1	.5	2.11	9	2.205
		1.0	2.78	9	3.270
		1.5	2.78	9	3.073
		Total	2.56	27	2.792
	0	.5	.78	9	1.641
		1.0	2.89	9	3.257
		1.5	4.22	9	3.563
		Total	2.63	27	3.176
	Total	.5	1.44	18	2.007
		1.0	2.83	18	3.167
		1.5	3.50	18	3.312
		Total	2.59	54	2.962
50	-1	.5	2.33	9	1.658
		1.0	1.89	9	2.315
		1.5	4.78	9	3.346
		Total	3.00	27	2.760
	0	.5	1.89	9	1.269
		1.0	1.78	9	1.986
		1.5	3.22	9	3.232
		Total	2.30	27	2.317
	Total	.5	2.11	18	1.451
		1.0	1.83	18	2.093
		1.5	4.00	18	3.290
		Total	2.65	54	2.549
Total	-1	.5	2.94	36	3.978
		1.0	3.47	36	4.960
		1.5	3.64	36	3.109
		Total	3.35	108	4.059
	0	.5	3.28	36	4.247
		1.0	3.78	36	3.658
		1.5	4.75	36	4.384
		Total	3.94	108	4.116
	Total	.5	3.11	72	4.089
		1.0	3.63	72	4.330
		1.5	4.19	72	3.815
		Total	3.64	216	4.088

There does not appear to be any consistent pattern in the effect of the item distribution. For the 20-item tests, the pool with a 1.0 SD seems to do worse than the pools with either larger or smaller SD's. For the 30-item tests and the 50-item tests, this pattern is reversed. For the 40-item tests, the smallest SD has the least difference in number passed and this number increases as the SD increases.

The last variable examined was the amount of overlap allowed between tests. The results are in Table 6.

Table 6. Pass Rate by Percent Overlap Within Test Length

Report

Difference in Number Who Pass				
Number of Items on Test	Percent Overlap	Mean	N	Std. Deviation
20	0	7.17	18	7.172
	20	5.44	18	4.422
	40	5.22	18	4.710
	Total	5.94	54	5.537
30	0	3.56	18	4.246
	20	2.72	18	3.707
	40	3.89	18	3.462
	Total	3.39	54	3.779
40	0	2.78	18	2.981
	20	2.00	18	2.890
	40	3.00	18	3.087
	Total	2.59	54	2.962
50	0	2.89	18	3.123
	20	2.11	18	1.568
	40	2.94	18	2.754
	Total	2.65	54	2.549
Total	0	4.10	72	4.937
	20	3.07	72	3.542
	40	3.76	72	3.625
	Total	3.64	216	4.088

When considered alone, the resulting pattern is also inconsistent. In general, the tests with 20% overlap had lower differences in pass numbers than the other tests and the tests with 0% overlap had the largest difference in the number passed.

Conclusions

Many organizations build relatively short tests to satisfy test constraints and equate them in efforts to be fair to the candidate. Despite these efforts, there can still be significant differences in the effective logit cut scores to which examinees are being held. This can have significant impacts on the pass/fail rates of alternative forms of an examination and the perceived fairness of the exam.

Factors that can affect the pass rate of these short examinations include:

- Test length: The shorter the test, the greater potential for different pass rates.
- The size of the available item pool: The larger the pool, the better. Effective pool size can be much smaller than the actual number of items in the pool due to enemy items.
- The distribution of item difficulties in the item pool that is used to construct the forms: Off-targeted item pools result in a decrease in the different number of individuals passed at the cost of reduced precision in measurement. An item pool that is spread out in a manner that is similar to the spread of the target population appears to work better than a distribution that is too concentrated or too spread out.

Some of this discrepant impact can be alleviated by the following actions:

- When equating test forms, carefully compare the score-to-measures and the effective logit cuts of the relevant forms.
- Attempt to sculpt the item pool so that the distribution of item difficulties is centered very near to the logit passing standard and there is an ample supply of items in each content domain.

- Whenever possible, increase the length of the test to reduce the probability of substantial pass/fail differences between forms.

References

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Appendix

Table A1. Tabulated Results for Test Length of 20 Items With 0% Percent Overlap

Pools	Cut score = 0			0%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
20	60	0	0.5	10	0.05	87.00	11	0.17	81.00
			1	10	0.13	84.00	10	0.03	90.00
			1.5	10	0.23	78.00	11	0.26	74.00
		-1	0.5	15	0.00	96.00	16	0.29	73.00
			1	14	0.02	92.00	15	0.30	71.00
			1.5	13	0.08	87.00	13	0.11	85.00
	100	0	0.5	10	0.02	92.00	11	0.19	81.00
			1	10	0.06	87.00	10	0.08	87.00
			1.5	10	0.13	84.00	11	0.21	80.00
		-1	0.5	14	0.16	82.00	14	0.18	81.00
			1	13	0.00	96.00	14	0.22	79.00
			1.5	12	0.11	85.00	12	0.23	78.00
	140	0	0.5	11	0.18	81.00	11	0.16	82.00
			1	10	0.02	92.00	11	0.21	80.00
			1.5	11	0.15	83.00	10	0.01	93.00
		-1	0.5	13	0.04	89.00	13	0.04	89.00
			1	12	0.06	87.00	12	0.05	87.00
			1.5	12	0.15	83.00	12	0.06	87.00

Table A2. Tabulated Results for Test Length of 20 Items With 20% Percent Overlap

Pools	Cut score = 0			20%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
20	60	0	0.5	10	0.06	87.00	11	0.17	81.00
			1	11	0.25	74.00	10	0.10	86.00
			1.5	9	0.06	87.00	10	0.00	96.00
		-1	0.5	15	0.01	93.00	15	0.01	93.00
			1	14	0.05	87.00	15	0.24	76.00
			1.5	14	0.30	71.00	14	0.25	74.00
	100	0	0.5	11	0.21	80.00	10	0.01	93.00
			1	10	0.08	87.00	10	0.02	92.00
			1.5	11	0.13	84.00	11	0.21	80.00
		-1	0.5	14	0.17	81.00	14	0.16	82.00
			1	13	0.01	93.00	14	0.20	81.00
			1.5	12	0.21	80.00	12	0.08	87.00
	140	0	0.5	11	0.16	82.00	11	0.18	81.00
			1	11	0.15	83.00	10	0.03	90.00
			1.5	10	0.10	86.00	10	0.03	90.00
		-1	0.5	13	0.05	87.00	13	0.04	89.00
			1	12	0.07	87.00	12	0.07	87.00
			1.5	12	0.17	81.00	12	0.16	82.00

Table A3. Tabulated Results for Test Length of 20 Items With 40% Percent Overlap

Pools	Cut score = 0			40%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
20	60	0	0.5	10	0.07	87.00	11	0.18	81.00
			1	10	0.07	87.00	10	0.00	96.00
			1.5	11	0.26	74.00	10	0.04	89.00
		-1	0.5	15	0.00	96.00	15	0.04	89.00
			1	14	0.05	87.00	14	0.07	87.00
			1.5	13	0.06	87.00	13	0.24	76.00
	100	0	0.5	10	0.01	93.00	11	0.21	80.00
			1	10	0.03	90.00	10	0.18	81.00
			1.5	10	0.03	90.00	11	0.11	85.00
		-1	0.5	14	0.16	82.00	14	0.19	81.00
			1	14	0.23	78.00	14	0.22	79.00
			1.5	12	0.18	81.00	12	0.17	81.00
	140	0	0.5	11	0.18	81.00	11	0.17	81.00
			1	11	0.19	81.00	11	0.18	81.00
			1.5	10	0.06	87.00	10	0.03	90.00
		-1	0.5	13	0.03	90.00	13	0.05	87.00
			1	12	0.02	92.00	12	0.07	87.00
			1.5	11	0.04	89.00	12	0.15	83.00

Table A4. Tabulated Results for Test Length of 30 Items With 0% Percent Overlap

Pools	Cut score = 0			0%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
30	90	0	0.5	16	0.03	90.00	16	0.11	85.00
			1	15	0.13	84.00	15	0.11	85.00
			1.5	16	0.21	80.00	14	0.00	96.00
		-1	0.5	24	0.15	83.00	24	0.13	84.00
			1	24	0.21	80.00	23	0.12	84.00
			1.5	19	0.15	83.00	20	0.07	87.00
	150	0	0.5	16	0.12	84.00	15	0.00	96.00
			1	15	0.07	87.00	15	0.02	92.00
			1.5	16	0.05	87.00	15	0.02	92.00
		-1	0.5	20	0.00	96.00	20	0.01	93.00
			1	19	0.10	86.00	19	0.10	86.00
			1.5	18	0.04	89.00	18	0.07	87.00
	210	0	0.5	15	0.00	96.00	15	0.00	96.00
			1	15	0.09	86.00	15	0.05	87.00
			1.5	16	0.06	87.00	16	0.09	86.00
		-1	0.5	19	0.05	87.00	19	0.03	90.00
			1	17	0.09	86.00	17	0.09	86.00
			1.5	17	0.10	86.00	17	0.08	87.00

Table A5. Tabulated Results for Test Length of 30 Items With 20% Percent Overlap

Pools	Cut score = 0			20%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
30	90	0	0.5	17	0.12	84.00	16	0.13	84.00
			1	15	0.16	82.00	16	0.08	87.00
			1.5	15	0.00	96.00	15	0.19	81.00
		-1	0.5	24	0.14	84.00	24	0.16	82.00
			1	24	0.18	81.00	23	0.13	84.00
			1.5	20	0.12	84.00	19	0.11	85.00
	150	0	0.5	16	0.13	84.00	16	0.13	84.00
			1	16	0.14	84.00	15	0.03	90.00
			1.5	16	0.07	87.00	16	0.05	87.00
		-1	0.5	20	0.02	92.00	20	0.00	96.00
			1	19	0.11	85.00	19	0.15	83.00
			1.5	18	0.14	84.00	19	0.12	84.00
	210	0	0.5	16	0.13	84.00	16	0.13	84.00
			1	15	0.03	90.00	15	0.03	90.00
			1.5	16	0.06	87.00	16	0.10	86.00
		-1	0.5	19	0.05	87.00	19	0.03	90.00
			1	17	0.13	84.00	17	0.11	85.00
			1.5	17	0.08	87.00	17	0.01	93.00

Table A6. Tabulated Results for Test Length of 30 Items With 40% Percent Overlap

Pools	Cut score = 0			40%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
30	90	0	0.5	15	0.02	92.00	17	0.13	84.00
			1	15	0.02	92.00	16	0.13	84.00
			1.5	14	0.10	86.00	16	0.18	81.00
		-1	0.5	24	0.14	84.00	24	0.15	83.00
			1	23	0.07	87.00	24	0.14	84.00
			1.5	20	0.11	85.00	19	0.17	81.00
	150	0	0.5	15	0.00	96.00	16	0.12	84.00
			1	15	0.04	89.00	15	0.00	96.00
			1.5	16	0.04	89.00	16	0.10	86.00
		-1	0.5	20	0.03	90.00	21	0.16	82.00
			1	19	0.10	86.00	19	0.11	85.00
			1.5	18	0.05	87.00	19	0.15	83.00
	210	0	0.5	15	0.01	93.00	15	0.01	93.00
			1	15	0.03	90.00	15	0.04	89.00
			1.5	16	0.06	87.00	16	0.06	87.00
		-1	0.5	19	0.05	87.00	19	0.03	90.00
			1	17	0.09	86.00	17	0.12	84.00
			1.5	17	0.11	85.00	17	0.11	85.00

Table A7. Tabulated Results for Test Length of 40 Items With 0% Percent Overlap

Pools	Cut score = 0			0%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
40	120	0	0.5	20	0.05	87.00	20	0.02	92.00
			1	20	0.06	87.00	20	0.06	87.00
			1.5	20	0.06	87.00	20	0.02	92.00
		-1	0.5	29	0.05	87.00	29	0.03	90.00
			1	29	0.01	93.00	29	0.07	87.00
			1.5	29	0.10	86.00	29	0.10	86.00
	200	0	0.5	21	0.08	87.00	21	0.09	86.00
			1	19	0.02	92.00	20	0.05	87.00
			1.5	21	0.00	96.00	22	0.06	87.00
		-1	0.5	26	0.04	89.00	26	0.03	90.00
			1	24	0.04	89.00	24	0.04	89.00
			1.5	24	0.04	89.00	24	0.03	90.00
	280	0	0.5	21	0.09	86.00	21	0.09	86.00
			1	21	0.09	86.00	21	0.08	87.00
			1.5	20	0.00	96.00	20	0.04	89.00
		-1	0.5	26	0.10	86.00	25	0.02	92.00
			1	23	0.07	87.00	23	0.05	87.00
			1.5	23	0.07	87.00	23	0.08	87.00

Table A8. Tabulated Results for Test Length of 40 Items With 20% Percent Overlap

Pools	Cut score = 0			20%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
40	120	0	0.5	20	0.03	90.00	20	0.03	90.00
			1	21	0.07	87.00	21	0.10	86.00
			1.5	20	0.08	87.00	20	0.05	87.00
		-1	0.5	29	0.03	90.00	29	0.03	90.00
			1	29	0.07	87.00	29	0.00	96.00
			1.5	29	0.05	87.00	28	0.02	92.00
	200	0	0.5	21	0.07	87.00	21	0.06	87.00
			1	20	0.04	89.00	19	0.01	93.00
			1.5	21	0.01	93.00	22	0.09	86.00
		-1	0.5	26	0.04	89.00	26	0.03	90.00
			1	24	0.05	87.00	24	0.06	87.00
			1.5	24	0.07	87.00	24	0.06	87.00
	280	0	0.5	21	0.09	86.00	21	0.09	86.00
			1	20	0.00	96.00	20	0.00	96.00
			1.5	20	0.05	87.00	21	0.07	87.00
		-1	0.5	26	0.10	86.00	25	0.03	90.00
			1	23	0.08	87.00	23	0.06	87.00
			1.5	23	0.06	87.00	22	0.02	92.00

Table A9. Tabulated Results for Test Length of 40 Items With 40% Percent Overlap

Pools	Cut score = 0			40%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
40	120	0	0.5	20	0.04	89.00	20	0.03	90.00
			1	21	0.04	89.00	20	0.01	93.00
			1.5	21	0.08	87.00	20	0.03	90.00
		-1	0.5	29	0.04	89.00	29	0.04	89.00
			1	29	0.02	92.00	29	0.06	87.00
			1.5	29	0.04	89.00	28	0.05	87.00
	200	0	0.5	21	0.06	87.00	21	0.06	87.00
			1	20	0.09	86.00	20	0.07	87.00
			1.5	22	0.01	93.00	22	0.09	86.00
		-1	0.5	26	0.02	92.00	26	0.02	92.00
			1	24	0.03	90.00	24	0.06	87.00
			1.5	23	0.00	96.00	24	0.08	87.00
	280	0	0.5	21	0.10	86.00	21	0.10	86.00
			1	20	0.00	96.00	21	0.10	86.00
			1.5	20	0.04	89.00	20	0.04	89.00
		-1	0.5	26	0.10	86.00	25	0.03	90.00
			1	23	0.08	87.00	23	0.04	89.00
			1.5	22	0.04	89.00	23	0.09	86.00

Table A10. Tabulated Results for Test Length of 50 Items With 0% Percent Overlap

Pools	Cut score = 0			0%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
50	150	0	0.5	25	0.07	87.00	25	0.07	87.00
			1	25	0.06	87.00	24	0.03	90.00
			1.5	25	0.00	96.00	23	0.10	86.00
		-1	0.5	38	0.00	96.00	38	0.01	93.00
			1	36	0.08	87.00	36	0.04	89.00
			1.5	37	0.07	87.00	34	0.00	96.00
	250	0	0.5	26	0.05	87.00	26	0.05	87.00
			1	25	0.04	89.00	25	0.03	90.00
			1.5	24	0.04	89.00	24	0.08	87.00
		-1	0.5	34	0.02	92.00	34	0.04	89.00
			1	30	0.03	90.00	30	0.04	89.00
			1.5	29	0.00	96.00	30	0.08	87.00
	350	0	0.5	25	0.03	90.00	25	0.02	92.00
			1	25	0.02	92.00	25	0.03	90.00
			1.5	26	0.06	87.00	26	0.04	89.00
		-1	0.5	32	0.02	92.00	32	0.01	93.00
			1	30	0.08	87.00	30	0.07	87.00
			1.5	28	0.04	89.00	28	0.06	87.00

Table A11. Tabulated Results for Test Length of 50 Items With 20% Percent Overlap

Pools	Cut score = 0	Mean	SD	20%					
				Test1				Test2	
TL	N			Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
50	150	0	0.5	25	0.04	89.00	24	0.02	92.00
			1	27	0.05	87.00	22	0.01	93.00
			1.5	23	0.05	87.00	24	0.09	86.00
		-1	0.5	38	0.01	93.00	38	0.00	96.00
			1	36	0.05	87.00	36	0.07	87.00
			1.5	37	0.06	87.00	35	0.12	84.00
	250	0	0.5	26	0.04	89.00	26	0.05	87.00
			1	25	0.03	90.00	25	0.04	89.00
			1.5	24	0.08	87.00	24	0.05	87.00
		-1	0.5	34	0.02	92.00	34	0.04	89.00
			1	30	0.03	90.00	30	0.01	93.00
			1.5	29	0.07	87.00	29	0.04	89.00
	350	0	0.5	25	0.03	90.00	25	0.02	92.00
			1	25	0.01	93.00	25	0.00	96.00
			1.5	25	0.03	90.00	26	0.07	87.00
		-1	0.5	32	0.02	92.00	32	0.02	92.00
			1	29	0.00	96.00	29	0.00	96.00
			1.5	28	0.03	90.00	28	0.06	87.00

Table A12. Tabulated Results for Test Length of 50 Items With 40% Percent Overlap

Pools	Cut score = 0			40%					
TL	N	Mean	SD	Test1			Test2		
				Raw Cut	Effective Logit Cut	#pass	Raw Cut	Effective Logit Cut	#pass
50	150	0	0.5	24	0.01	93.00	25	0.04	89.00
			1	26	0.02	92.00	23	0.02	92.00
			1.5	23	0.03	90.00	25	0.07	87.00
		-1	0.5	38	0.01	93.00	38	0.00	96.00
			1	36	0.05	87.00	36	0.07	87.00
			1.5	35	0.01	93.00	37	0.14	84.00
	250	0	0.5	26	0.04	89.00	26	0.05	87.00
			1	25	0.07	87.00	25	0.05	87.00
			1.5	23	0.01	93.00	24	0.02	92.00
		-1	0.5	34	0.02	92.00	34	0.05	87.00
			1	30	0.06	87.00	30	0.02	92.00
			1.5	29	0.03	90.00	29	0.04	89.00
	350	0	0.5	25	0.03	90.00	25	0.02	92.00
			1	25	0.02	92.00	25	0.02	92.00
			1.5	26	0.04	89.00	25	0.00	96.00
		-1	0.5	32	0.02	92.00	32	0.02	92.00
			1	29	0.01	93.00	30	0.07	87.00
			1.5	27	0.02	92.00	28	0.08	87.00