

**Item Response Theory Parameterization
of the Multistate Bar Exam**

Nathan A. Thompson and David J. Weiss

University of Minnesota

**Final Report
February 2006**

2005 Joe E. Covington Award for Bar Exam Research

Abstract

The Multistate Bar Examination (MBE) is a 200-item test with six content areas. The applicability of IRT calibration and scoring of the MBE was investigated. Calibration methods referenced either the overall full scale or the individual content areas. Three methods of IRT ability (θ) estimation were examined: maximum likelihood, maximum *a posteriori*, and expected *a posteriori*. It was found that there was little effect on θ estimates from different calibration methods, while the only difference in estimation methods was a slightly different standard deviation in θ estimates. Neither variable had much effect on correlations between θ estimates and the current method of scoring, number correct. It was concluded that the application of IRT was viable, and if applied, should be done with full-scale referenced calibration.

Item Response Theory Parameterization of the Multistate Bar Exam

The Multistate Bar Exam (MBE), developed by the National Conference of Bar Examiners (NCBE), is a 200-item multiple-choice test that is designed to assess each candidate's skill in applying basic legal principles in fact-based situations. The content of the questions is drawn from six areas: Constitutional Law, Contracts, Criminal Law, Evidence, Real Property, and Torts. The exam is scored by counting the number of correct answers for each examinee, and these scores are converted to scaled scores. The test is administered twice a year, with a new form for each test date that is equated to the scale, which remains constant, using the three-parameter item response theory (IRT) model. The content areas are not scored separately; their purpose is to ensure that examinees master more than one topic.

This study compared several methods of IRT parameterization for the MBE. The three-parameter logistic model was first fitted to the item responses, and then the item parameter estimates were used to obtain ability (θ) estimates for each examinee. Three θ estimation methods were used, which were compared to each other, as well as to number correct (NC) scoring. The standard error of measurement (SEM) accompanying each examinee's θ estimate was also calculated, as well as an index of the examinee's fit to the three-parameter logistic model.

When content areas are present in a test, there are three approaches for scoring, regardless of whether NC or IRT scoring is employed. The simplest approach is to score the examinee on the full test only, which is the current practice with the MBE. A second approach is to score the content areas individually. This can be used to ensure that the examinee meets a minimum standard in each area. Lastly, the previous two can be combined, and the examinee is evaluated both on the content areas and the overall score.

When IRT scoring is used, this paradigm must also be applied to the item calibration procedure, in addition to examinee scoring. If the test is assumed to have only one main dimension, and therefore has only one overall score, all items in the test should be parameterized along this single dimension. If the content areas are considered distinct, the items should be parameterized in each content area separately. The reason for this paradigm is the fact that the most frequently used IRT models assume a unidimensional trait. If the content areas are distinct, then the overall score is not unidimensional—only the content area scores are. Both item parameters and examinee scores should reflect these differences, if present.

These two approaches, full scale (FS) and content area (CA), are easily achieved with conventional IRT calibration methods. The comparison of the two approaches via IRT calibration as a method of assessing unidimensionality was first suggested by Bejar, Weiss, and Kingsbury (1977), elucidated by Bejar (1980), and empirically applied by Bejar (1988). It was used in the investigation of item banks for the CAT-ASVAB by Segall, Moreno, and Hetter (1997), who termed them *unidimensional treatment* and *pool splitting*, respectively.

In a third approach, items are calibrated simultaneously on the content areas and the overall dimension if both are used to evaluate the examinee. This is accomplished via the bifactor model, a factor analysis model that assumes all items correlate with a main dimension *and* with moderately distinct component content area dimensions (Gibbons & Hedeker, 1992). These correlations, or factor loadings, can then be converted into IRT item parameters. The bifactor model was not applied in this study because the substantial comparability of FS and CA calibrations suggested that bifactor calibrations would not produce noticeably different results.

Once item parameters were estimated, they were used in conjunction with the examinee's item responses to estimate the likelihood function of θ . The likelihood function represents the likelihood of the examinee's responses as a function of θ and the item parameters. The likelihood function was then used to obtain a single estimate of θ for the examinee. This was done via three methods: maximum likelihood estimation (MLE), expected *a posteriori* (EAP) and maximum *a posteriori* (MAP). These, as well as the IRT calibrations, are discussed in detail below.

In order to evaluate differences between the calibration and scoring methods, all were used to estimate each θ for each examinee. For instance, the full θ was estimated not only with the FS item parameters, but also with the CA parameters, and both of these were used with all three scoring methods. This was then repeated for each of the content areas, which produced 42 θ estimates for each examinee by crossing score (1 full, 6 content areas) by calibration (FS, CA) by scoring method (MLE, MAP, EAP). Each of these 42 θ estimates was accompanied by an SEM estimate and an index of person fit. Note that the calibration methods are named with the acronyms of FS and CA, while the scores (θ estimates) are referred to as full and area.

The purpose of this study was to evaluate the appropriateness of each of these methods for scoring the MBE. The more that the different scores and calibration methods produced θ estimates that correlated highly with each other, the more a unidimensional approach is justified. If content area scores do not always correlate highly with each other or the main dimension, or if CA calibration leads to substantially different θ estimates than FS calibration, then a content area based approach to scoring the MBE is likely more valid. Such a conclusion was further investigated through a factor analysis of the data.

Method

Data

The data used were from the July 1998 administration of the MBE, and consisted of 43,538 examinees from across the United States. All 200 items were four-alternative multiple-choice items, scored as correct or incorrect. Items were distributed evenly across the six content areas. Contracts and Torts each had 34 items, while the other four areas each had 33 items.

Factor Analysis

Before calibrating the data with IRT, a principal factor analysis was performed on the responses to all 200 items to assess the dimensionality of the data. The item-level analysis was done with MicroFACT (Assessment Systems Corporation, 1997), which uses tetrachoric correlations. Squared multiple correlations were used as the initial communalities estimates, with a maximum of 200 iterations for the estimation of final communalities. The ϵ for communality estimation and rotation was 0.0001. A six-factor solution was obtained to determine if the six content areas fell into six factors. Factors were rotated with varimax rotation.

A principal factor analysis was also performed on the NC scores from each content area with SPSS. This analysis used the matrix of Pearson correlations between the content areas. Loadings were estimated with maximum likelihood, and it was also rotated with varimax rotation.

IRT Model

IRT assumes that the probability of responding to an item can be modeled by a logistic function. The probability P_i of an examinee correctly responding to item i is (Hambleton & Swaminathan, 1985, Eq. 3.3),

$$P_i(X = 1 | \theta_j) = c_i + (1 - c_i) \frac{\exp[Da_i(\theta - b_i)]}{1 + \exp[Da_i(\theta - b_i)]}, \quad (1)$$

where $P_i(X = 1 | \theta_j)$ is the probability of examinee j correctly responding to item I , D is a scaling constant equal to 1.702, and e is approximately 2.71. The item parameter a_i is the item discrimination and is related to the maximum slope of the item response function (IRF); b_i is the difficulty, or location, of the IRF, representing the point on θ where an examinee with that level has a probability of $0.5 + (c_i / 2)$ of correctly responding to the item; and c_i is a pseudoguessing parameter representing the lower asymptote of the IRF. The IRF models the response to an item as a function of item and person parameters.

One goal of IRT is to estimate the values of these parameters for each item. If the item fits the model, then Equation 1 describes the function of the item. More importantly, the parameters can be used to determine a likelihood function for estimating examinee θ . This likelihood function (Embretson & Reise, 2000, p. 161),

$$L(x_{j1}, x_{j2}, \dots, x_{ji} | \theta_j, a_i, b_i, c_i) = P_i(\theta_j)^{x_{ji}} Q_i(\theta_j)^{1-x_{ji}}, \quad (2)$$

represents the likelihood of the examinee's responses given each possible level of θ , where x_{ji} is 1 if the item is answered correctly and 0 otherwise, and $Q_j = 1 - P_j$. If the likelihood of the observed responses is large for a given θ value, it is more likely that the θ value would yield the observed pattern of item responses. The three methods of scoring used in this study (described below) differ in how they define that last step of determining the greatest likelihood.

IRT Calibration

IRT parameters for the three-parameter logistic model (Equation 1) were estimated for each item. The calibration procedure was referenced to either the full scale or the content areas, producing two sets of a , b , and c parameter estimates. Thus, there were seven sets of calibrations. The first calibration included all 200 items, placing all item parameters on the same scale. The remaining six calibrations placed the item parameters from each content area on a scale referenced only to the other items in that content area. Once the item parameters were estimated separately for each content area, they were combined to produce a single set for all 200 items. This was done by simply concatenating the item parameter estimates for each set of 33 or 34 items in each area.

The calibration software used was XCALIBRE (Assessment Systems Corporation, 1997), which implements marginal maximum likelihood estimation of item parameters. Default "floating" priors were used for the item calibration, with specified prior distributions of $a \sim N(0.75, 0.12)$, $b \sim N(0, 1)$, and $c \sim N(0.25, 0.025)$. A floating prior means that the prior distribution moments are updated at the end of each iteration of the estimation algorithm, to provide a new set of priors for the next iteration. The maximum number of iterations for the calibration was fixed at 100. If the parameter estimates of an item did not converge after 100 iterations, the item was eliminated and item parameters were estimated for the remaining items.

Some items did not fit the three-parameter model. These items are listed in Table 1. For the FS calibration with all 200 items, eleven items had parameter estimates that did not converge in the

estimation process. The eleven items are marked in the first column of Table 1 with an “x.” These were removed, and the calibration repeated. This time, four items had parameter estimates that did not converge; these are indicated in the second column of Table 1. These were also removed, and the calibration done for a third time, with no nonconverging items. The parameters from the third calibration of the remaining 185 items constituted the set of item parameters used for the FS analyses.

Table 1: Nonconvergent Items Deleted by Full Scale (FS) and Content Area (CA) IRT Calibrations

Item	Calibration		
	FS, 200 Items	FS, 189 Items	CA, 200 Items
9			x
11			x
23	x		x
31		x	x
37	x		
42			x
46	x		
49			x
52			x
58	x		x
60			x
74			
78	x		
89		x	
92	x		
105			x
123	x		x
125		x	x
137		x	
175	x		
180	x		
191			x
197			x
198	x		
199	x		

The same process was applied to the CA calibrations. The first round of six calibration runs of XCALIBRE led to fourteen items being identified as nonconvergent, four of which were also identified as nonconvergent in the FS calibration (see Table 1). These fourteen were eliminated from the analysis, and the remaining 186 items calibrated in a second round of six runs. All 186 converged during this round, so the set of CA item parameters contained 186 items. Out of the 14 nonconvergent CA items, five were also nonconvergent with FS.

Most of the items that did not converge did so because the parameter estimates were outside of normal ranges for IRT parameters. For instance, on the final iteration of the first FS calibration with all 200 items, item 23 had $a = 0.53$, $b = -3.00$, and $c = 0.06$. The item was extremely easy for most examinees, but those with low θ had little chance of guessing the correct answer. This can be contrasted with item 175 ($a = 0.25$, $b = 3.00$, and $c = 0.81$) and item 198 ($a = 0.20$, $b = 2.16$, and $c = 0.95$). These

items were answered correctly by most examinees (classical difficulties of 0.85 and 0.98, respectively), leading to extremely high pseudoguessing parameters. Even an examinee with low θ had a very high probability of correctly answering those two items. This, however, changed the shape of the rest of the IRF, preventing an accurate estimation of the function's location (the b parameter), hence the relatively high difficulty (b) parameters, even though the items were correctly answered by the vast majority of the sample.

These results might be partly due to the criterion-referenced construction of the MBE. Items are not chosen because of their fit to an IRT model or their discriminations. The topic of an item could be known by every examinee to be on the test, yet is still included because it covers information that every practicing lawyer should know. Items 175 and 198 could be such items.

Estimation of θ

Once the likelihood function is estimated for each examinee using their item responses and a set of IRT parameters, a single θ estimate must be obtained. This can be formulated in three ways. The first method, maximum likelihood estimation (MLE), does not assume a prior distribution of θ , and estimates θ for each examinee as the level of θ at which the likelihood function reaches its maximum. The other two methods make an assumption regarding the prior distribution of θ , which was $N(0,1)$ in this study, and are referred to as Bayesian methods. The prior distribution is a starting point for the likelihood function, so that all successive estimates are multiplied by the normal curve. The first of these, maximum *a posteriori* (MAP) is similar to maximum likelihood in that it evaluates the maximum of the likelihood function. It differs from MLE only in that the MLE likelihood function is multiplied by the prior distribution, producing what is known as the posterior distribution. The third method evaluates the mean of the posterior distribution, and is called the expected *a posteriori* (EAP) estimator.

For each θ estimation method, the observed SEM was also calculated. This is an index of the precision of the θ estimate, and is calculated from the second derivative of the likelihood function, which is related to its standard deviation (Thissen & Wainer, 2001, p. 102):

$$SEM(\theta) = \sqrt{-1 / E\left(\frac{\partial^2 L}{\partial \theta^2}\right)}. \quad (3)$$

The SEM is an index of the precision of measurement; the smaller the SEM, the more precise the measurement of the examinee. Therefore, if the item parameters from one calibration produce smaller SEMs across the sample than the other calibration, the first calibration method produces more precise measurements.

Person fit is a method of assessing how well the responses of examinees are accounted for by the model. The most commonly used measure of IRT person fit is Z_L (Embretson & Reise, 2000, Eq. 9.13; Reise, 1990), more often called L_z ,

$$L_z | \theta_j = \sum [LogL | \theta_j - \Sigma E(LogL | \theta_j)] / (\Sigma V(LogL | \theta_j))^{1/2}, \quad (4)$$

where (Embretson & Reise, 2000, Eq. 9.12),

$$V(LogL | \theta_j) = \Sigma (P_i | \theta_j)(Q_i | \theta_j) [\log(P_i | \theta_j) / (Q_i | \theta_j)]^2. \quad (5)$$

and (Embretson & Reise, 2000, Eq. 9.11),

$$E(\text{Log}L | \theta_j) = \sum \{(P_i | \theta_j) \times \log(P_i | \theta_j) + (Q_i | \theta_j) \times \log(Q_i | \theta_j)\} \quad (6)$$

This is calculated for each person j that responds to each item i , with a likelihood function L . L_z is readily interpretable in both directions. A positive value means that the model fits an individual's response pattern better than would be expected from the probabilistic IRT model, while a negative value indicates worse fit. It is referenced against the standard normal distribution as the null distribution, so values of L_z have meaning similar to standard scores. A value greater than 1.96 in either direction indicates statistical significance, with a Type I error equal to approximately .05. The fit of a model for a group of examinees can be evaluated by determining the number of examinees that do not fit, or have a value of L_z less than -1.96.

The two sets of item parameters (FS 185 and CA 186) were used to obtain θ estimates for the full scale and for each of the six content areas. This resulted in 14 θ estimates for each examinee. The three methods of θ estimation, MLE, MAP, and EAP, were used for each of these scores, producing 42 θ estimates. The SEM and L_z were also calculated for each θ estimate. This was all done with a recent revision of the program SCOREALL (Assessment Systems Corporation, 2006).

Results

Factor Analysis

At the item level, six factors adequately reproduced the item intercorrelation matrix, as the mean residual was -0.0001, the root mean square of the residuals was 0.01610, and the goodness-of-fit index was 0.9818. The residual is the difference between the actual values in the item intercorrelation matrix and the values predicted by the six-factor model. A mean residual and root mean square residual near 0.0 indicate good fit, as does a goodness-of-fit index near 1.0. Thirty-two iterations were needed for final communality estimates.

The eigenvalues, or amount of variance explained by each of the six factors, are presented in Table 2 (the rotated and unrotated factor loading matrices are presented in Appendix A, as well as the estimated communalities). The unrotated solution was strongly unidimensional, supporting a FS calibration approach. The varimax rotated solution was not as strongly unidimensional, but still indicated that each of the content areas did not represent a distinct factor; an examination of the factor loadings of the individual items did not discover any pattern related to the content of the items.

Table 2: Eigenvalues from the First Six Factors for Item-Level Factor Analysis (IL) and Factor Analysis of Content Area Scores (CA)

Factor	Eigenvalues		
	IL Unrotated	IL Rotated	CA Unrotated
1	18.57	10.86	3.71
2	3.34	7.75	0.55
3	1.81	3.31	0.50
4	1.57	2.79	0.45
5	1.20	1.63	0.41
6	1.06	1.20	0.38

Unidimensionality was also supported by a factor analysis of NC scores for each content area. The unrotated eigenvalues are presented in the third column of Table 2, with the first factor accounting for approximately 61.89% of the common variance. The same was true with a random split of the sample into two halves: the first unrotated factor accounted for 61.91% of the common variance in one half and 61.86% in the other half. SPSS extracted only that single factor in the rotated solution. The six content areas all loaded strongly on the extracted rotated factor (Table 3) and the NC scores correlated positively (between .48 and .61) with each other (Table 4).

Table 3: Loadings of Content Area NC Scores on Factor 1

<i>Content Area</i>	<i>Loading</i>
Constitutional Law	0.737
Contracts	0.762
Criminal Law	0.708
Evidence	0.772
Real Property	0.781
Torts	0.658

Table 4: Correlation Matrix of Content Area NC Scores

<i>Content Area</i>	Constitutional Law	Contracts	Criminal Law	Evidence	Real Property	Torts
Constitutional Law						
Contracts	0.559					
Criminal Law	0.514	0.525				
Evidence	0.572	0.577	0.566			
Real Property	0.582	0.613	0.544	0.596		
Torts	0.481	0.509	0.484	0.507	0.498	

Reliability

The internal consistency reliability of the whole test, as well as the content area subscales, was calculated during the IRT calibration. The KR-20 coefficient for the full scale of 185 items, after the deletion of non-fitting items, was 0.895. The KR-20 coefficients for the individual content areas were (with the number of fitting items in parentheses): Constitutional Law, 0.610 (29); Contracts, 0.591 (29); Criminal Law, 0.575 (33); Evidence, 0.640 (30); Real Property, 0.648 (33); Torts, 0.439 (32). Given the numbers of items in the content area subscales, content area scores had relatively low reliabilities that are not high enough for individual measurement by contemporary standards.

Item Parameters

Before the item parameters were used to estimate θ , the parameters themselves were examined. If the content areas are distinct enough that they should be considered separately, the FS and CA calibrations would produce different item parameters, which would in turn lead to different θ estimates. The item parameters are listed in Appendix B.

Bejar (1988) suggested that a scatterplot of item difficulty parameters from the two calibration methods be examined. When the data points deviate from a 45-degree line with a slope of 1.0 and intercept of 0.0, the assumption of unidimensionality is supported less strongly. Figure 1 presents this scatterplot for the MBE *b* parameters. The data fell closely along a 45-degree line, but the slope was slightly greater than 1.0. A linear regression of the parameters estimated a slope of 1.07 and intercept of -0.29. This suggests that the MBE is moderately unidimensional, using Bejar's criterion. Figures 2 and 3 present scatterplots of the *a* and *c* parameters, respectively. A regression of the *a* parameters estimated a slope of 0.94 and an intercept of -0.03, while a regression of the *c* parameters estimated a slope of 0.69 and an intercept of 0.02.

Figure 1: Scatterplot of Item Difficulty Parameter Estimates for FS and CA Calibrations

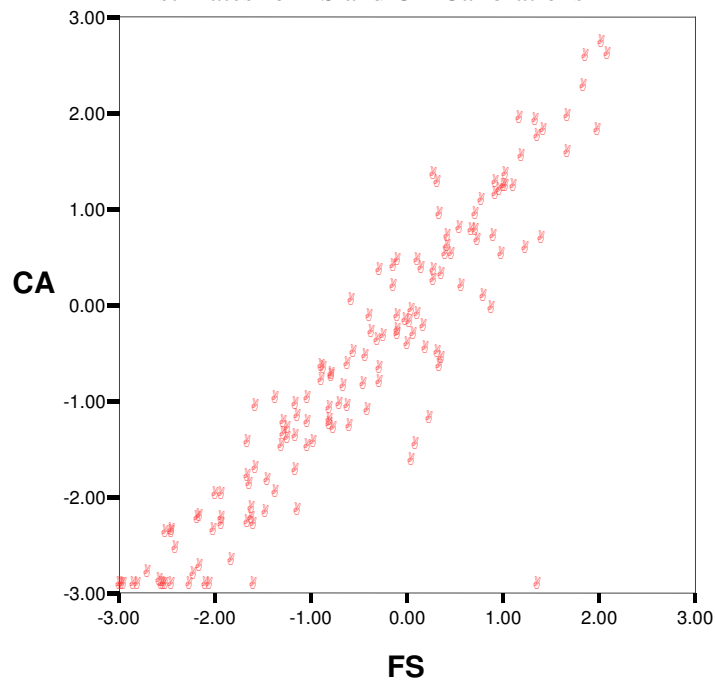


Figure 2: Scatterplot of Item Discrimination Parameter Estimates for FS and CA Calibrations

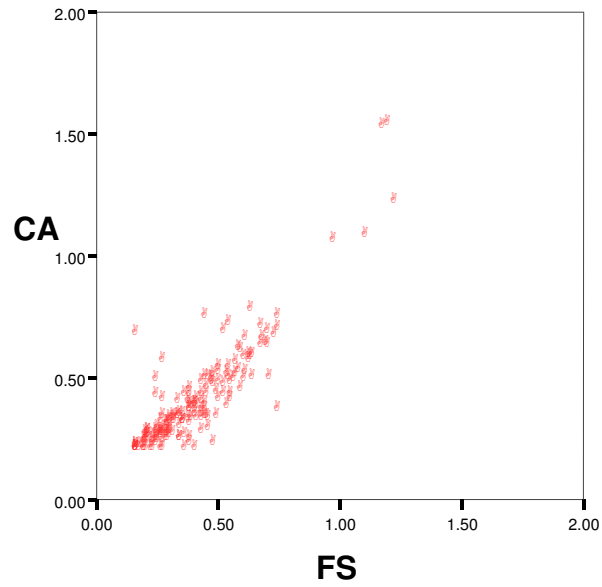
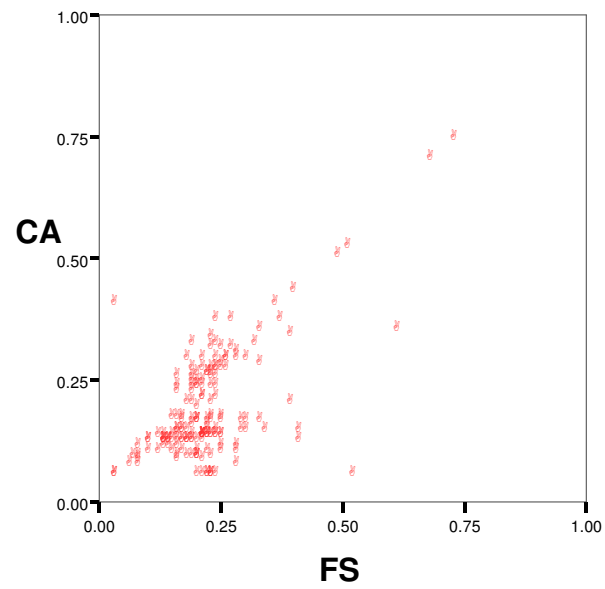


Figure 3: Scatterplot of Item Pseudoguessing Parameter Estimates for FS and CA Calibrations



The item parameters from the FS and CA calibrations correlated 0.88 for the a parameter, 0.95 for the b parameter, and 0.62 for the c parameter. These results support the equivalence of the FS and CA calibrations. The item parameters produced by the two calibration methods were very similar, as the correlations were quite high for a and b , and moderately high for c . Moreover, differences in c might be due to the difficulties encountered in estimating c for many items rather than actual differences in the parameter.

The descriptive statistics, including means and standard deviations (SDs) of a , b , and c for the FS and CA calibration are shown in Table 5. The descriptive statistics of the item parameters for each content area separately are shown in Appendix C. FS calibration produced higher mean parameter values for all three parameters. Note the SDs of the b parameters; if item difficulties were distributed similar to examinee θ , the SDs would be near 1.0. However, they were nearly double that, as the result of more items at the extremes of the scale than a normal distribution, especially the low difficulty end of the θ continuum. This can be seen in the grouped frequency distributions of the a , b , and c parameters in Figures 4, 5, and 6, respectively.

Table 5: Item Parameter Descriptive Statistics

	Parameter	N	Minimum	Maximum	Mean	SD	Skewness	Kurtosis
FS Calibration	a	185	0.20	1.26	0.45	0.19	1.53	3.73
	b	185	-3.00	3.00	0.03	1.77	0.16	-0.93
	c	185	0.05	0.75	0.24	0.10	1.79	6.26
CA Calibration	a	186	0.20	1.53	0.39	0.21	2.61	10.08
	b	186	-2.99	3.00	-0.31	1.96	0.32	-1.03
	c	186	0.05	0.95	0.19	0.14	2.41	8.39

Figure 4: Grouped Frequency Distribution of a Parameters for FS and CA Calibrations

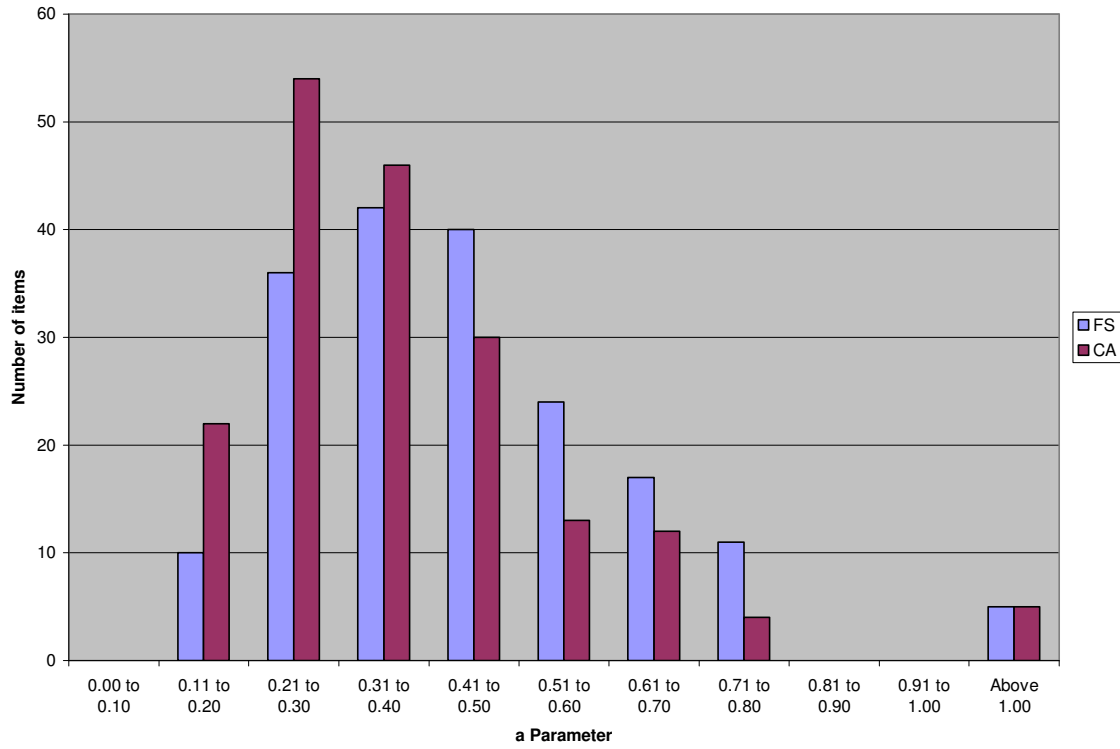


Figure 5: Grouped Frequency Distribution of b Parameters for FS and CA Calibrations

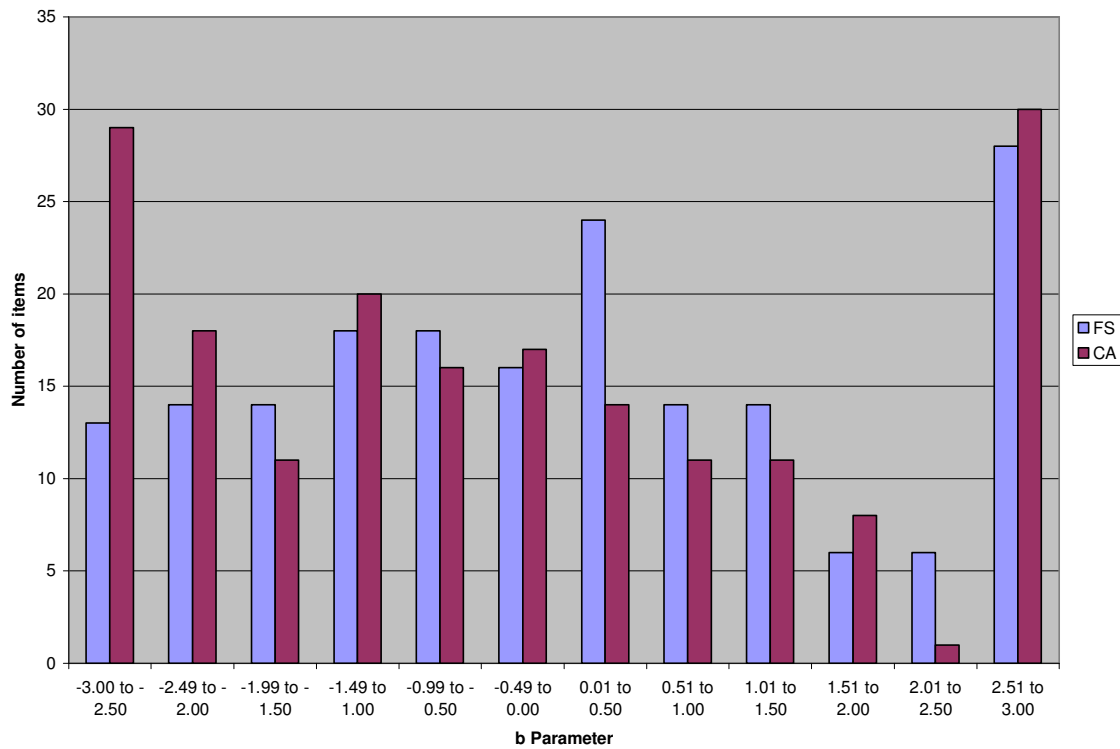
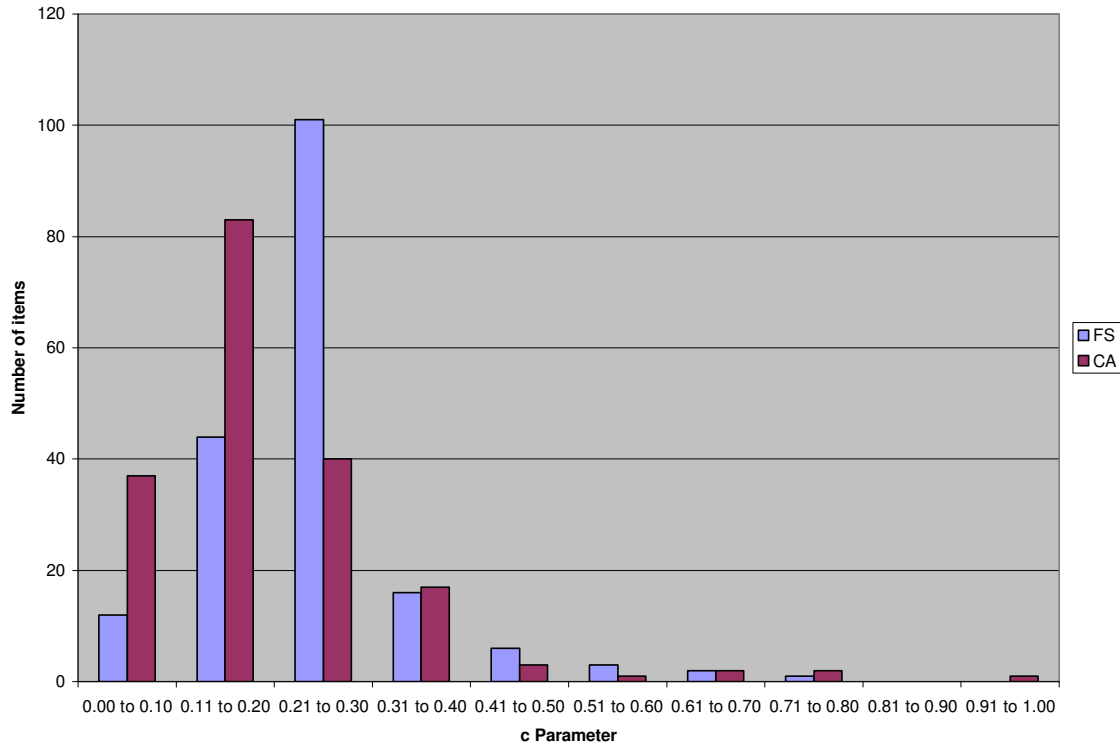


Figure 6: Grouped Frequency Distribution of c Parameters for FS and CA Calibrations



Person Parameters

The correlations among the θ estimates from the three θ estimation methods and two calibration methods for the full scale are presented in Table 6. Also shown are correlations with the NC scores based on the 185 items used in the FS calibration and those based on all 200 items. The lowest value in this table is 0.977, indicating that the ordering of examinee scores was the same regardless of scoring procedure. Scoring examinees with IRT rather than NC did not substantially change the test scores at the person level.

Table 6: Correlation Matrix of Full θ Estimates Using FS and CA Parameter Estimates, and NC Scores

	MAP FS	MLE FS	EAP FS	MAP CA	MLE CA	EAP CA	NC185	NC200
MAP FS								
MLE FS	0.998							
EAP FS	1.000	0.998						
MAP CA	0.990	0.990	0.989					
MLE CA	0.987	0.989	0.987	0.999				
EAP CA	0.991	0.990	0.990	1.000	0.999			
NC185	0.989	0.986	0.988	0.983	0.980	0.984		
NC200	0.986	0.983	0.985	0.981	0.977	0.981	0.997	

Within the IRT methods, the two Bayesian approaches, MAP and EAP, performed equivalently, correlating 0.997 for the FS calibration. They also had similar means and SDs, with MAP having a mean of 0.01 and SD of 0.97, and EAP a mean of 0.00 and SD of 0.97 (Table 7). These were similar due to the use of the same prior distribution, which also tended to draw the θ estimates closer to 0.00. MLE estimates did not have this trend, so while MLE had a similar mean (0.00) for the FS calibration, it had a slightly larger SD (1.09). This was the only noticeable difference among the IRT scoring methods in terms of θ estimation for the full scale.

Table 7: Descriptive Statistics for Full-Scale θ Estimates
Based on FS and CA Item Calibrations

	Estimation and Calibration	Minimum	Maximum	Mean	SD
θ Estimates	MAP, CA	-4.04	3.08	0.01	0.97
	EAP, CA	-3.49	3.08	0.00	0.97
	MLE, CA	-5.79	3.39	0.00	1.09
	MAP, FS	-4.30	3.13	0.00	0.91
	EAP, FS	-3.58	3.12	-0.01	0.92
	MLE, FS	-5.99	3.43	-0.01	1.03
SEM	MAP, CA	.26	.54	0.30	0.02
	EAP, CA	.26	.54	0.30	0.03
	MLE, CA	.27	.55	0.32	0.05
	MAP, FS	.25	.50	0.29	0.03
	EAP, FS	.26	.51	0.30	0.03
	MLE, FS	.26	.51	0.31	0.05
Person Fit	MAP, CA	-5.32	3.63	0.14	0.95
	EAP, CA	-5.30	3.65	0.16	0.95
	MLE, CA	-5.30	3.71	0.09	0.94
	MAP, FS	-5.15	3.28	-0.04	0.95
	EAP, FS	-5.13	3.31	-0.03	0.95
	MLE, FS	-5.13	3.34	-0.10	0.93

Table 8 shows the same correlation matrix as Table 6, but for θ estimates from the Constitutional Law content area. The correlations were slightly lower (ranging from 0.938 to 0.995), but this might be due to the fact that there was one-sixth as many items used in θ estimation. Nevertheless, it can still be seen that the various methods produced highly similar scores—the same result that was found with full-scale scores. Correlation matrices for each of the other content areas (in Appendix D) were similar. Scores on the content areas correlated moderately with each other and the full-scale θ estimates (Table 9), though the correlations with the full scale were somewhat higher. However, this is at least partially due to the fact that the responses used to estimate θ in each content area were also used to estimate the full-scale θ .

Table 8: Correlation Matrix of θ Estimates for Constitutional Law

	MAP CA	MLE CA	EAP CA	MAP FS	MLE FS	EAP FS	NC CA
MAP CA							
MLE CA	0.982						
EAP CA	0.995	0.983					
MAP FS	0.964	0.948	0.962				
MLE FS	0.954	0.955	0.955	0.990			
EAP FS	0.957	0.947	0.963	0.994	0.990		
NC CA	0.953	0.938	0.949	0.961	0.953	0.954	

Table 9: Correlation of FS θ Estimates for All Content Areas and Full Scale

	Torts	Real Property	Evidence	Criminal Law	Contracts	Constitutional Law	Full Scale
Torts							
Real Property	0.560						
Evidence	0.560	0.618					
Criminal Law	0.505	0.540	0.549				
Contracts	0.535	0.600	0.569	0.497			
Constitutional Law	0.547	0.623	0.608	0.507	0.563		
Full Scale	0.743	0.837	0.820	0.728	0.768	0.817	

Table 10 presents the same descriptive statistics for estimates as Table 7, but for the Constitutional Law content area. The tables for the other five content areas are in Appendix E. For Constitutional Law, the SD of the θ estimates were much larger with MLE than the Bayesian methods. The θ estimate means were all near 0.00. Similar results were found for Criminal Law, Evidence, and Real Property. The mean θ estimates in Contracts were slightly lower with CA calibration, while they were slightly higher in Torts.

Table 10: Descriptive Statistics for Constitutional Law Estimates

	Calibration and Estimation	Minimum	Maximum	Mean	SD
θ Estimates	MAP, CA	-3.05	2.21	0.04	0.77
	EAP, CA	-2.97	2.07	0.03	0.79
	MLE, CA	-5.99	5.89	0.04	1.37
	MAP, FS	-3.09	2.24	0.03	0.81
	EAP, FS	-2.59	2.07	0.02	0.82
	MLE, FS	-5.97	5.77	0.07	1.40
SEM	MAP, CA	.49	.83	0.57	0.04
	EAP, CA	.51	1.86	0.59	0.05
	MLE, CA	.56	8.80	0.79	0.25
	MAP, FS	.51	.72	0.59	0.03
	EAP, FS	.52	2.31	0.60	0.04
	MLE, FS	.59	4.73	0.79	0.18
Person Fit	MAP, CA	-4.28	2.12	0.04	0.86
	EAP, CA	-4.21	2.15	0.06	0.84
	MLE, CA	-4.26	2.58	-0.04	0.74
	MAP, FS	-4.55	2.32	0.16	0.78
	EAP, FS	-3.89	2.47	0.19	0.79
	MLE, FS	-4.48	2.76	0.09	0.71

SEMs and Person Fit

The descriptive statistics of the SEMs for the full-scale θ estimates are also in Table 7. The SEM mean and SD for the two Bayesian methods were slightly lower than the SEM mean and SD for MLE, though the difference was too small – 0.01 or 0.02 – to be of significance. There was a similar difference between the mean SEMs of the two calibration methods. The SDs of the SEMs were very small (0.002 to 0.005), indicating that the precision of measurement varied little across examinees.

There was a much more noticeable effect found with person fit of the full-scale scores. While there was little difference between the calibration methods in terms of the SD of person fit statistics, there was a difference between the means. All three CA means were low positive, while all three FS means were slightly negative. This means that examinees fit the model slightly better when CA calibrated item parameters were used as opposed to FS calibrated parameters. Nevertheless, the difference was not large.

Table 10 shows the same statistics for the first content area, Constitutional Law. Again, MLE had higher mean SEMs (0.79) than the Bayesian methods (0.57 to 0.60) due to the Bayesian priors. This difference was much more pronounced than with the full-scale scores, which had a difference of only about 0.01 to 0.02 for each θ estimation method. All three estimation methods had larger mean SEMs than with the full-scale scores, which is due to there only being one-sixth as many items, and therefore one-sixth as much information. A comparable pattern was found in the other content areas (Appendix E).

MLE estimation and CA calibration had slightly lower mean levels of person fit, but this difference was not large. This was opposite the full-scale scores, where CA had slightly better person fit. Again, a comparable pattern was found for the other five content areas (Appendix E). For most content areas and scoring methods, there was a tendency of examinees to overfit the IRT model, as indicated by positive nonzero mean L_Z values.

Table 11 gives the proportion of examinees with an L_Z statistic less than -1.96 for each instance of θ estimation. Values in that range indicate that examinees are not fitting the model, with a probability of 0.025. Fewer examinees had misfit with content area scores than with the full-scale score. More importantly, however, is that FS calibrations had fewer examinees with misfit than CA for the full-scale scores. FS calibration also produced proportions of misfit proportions lower than CA calibration for 11 of the 18 content area θ estimates (Appendix E), crossing the six content areas by three θ estimation methods.

Table 11: Proportion of Examinees with $L_Z < -1.96$

	<i>Full score</i>	<i>Full</i>	<i>Constitutional Law</i>	<i>Contracts</i>	<i>Criminal Law</i>	<i>Evidence</i>	<i>Real Property</i>	<i>Torts</i>
FS	MLE	0.020	0.005	0.017	0.014	0.016	0.022	0.013
	MAP	0.020	0.009	0.019	0.018	0.015	0.017	0.017
	EAP	0.019	0.007	0.017	0.017	0.014	0.016	0.016
CA	MLE	0.031	0.013	0.021	0.014	0.015	0.020	0.016
	MAP	0.032	0.020	0.023	0.017	0.017	0.017	0.024
	EAP	0.030	0.019	0.022	0.017	0.017	0.016	0.018

Discussion and Conclusions

The most important result of this study was the extremely high correlations among the IRT θ estimates, and between the θ estimates and the NC scores. IRT scoring did not order examinees any differently than NC scoring. Given the greater complexity of IRT, as well as the problems with items fitting the model, it appears that IRT scoring of the MBE adds little to the process of ordering examinees. However, IRT scoring offers IRT-based alternatives for linking and equating (and, therefore, item banking), as well as providing individual SEMs, measures of person fit, and explicit information criteria for test assembly. IRT equating is already employed in the development of the MBE.

The MBE test, while not completely unidimensional, is unidimensional enough that there was virtually no difference between θ estimates using FS and CA calibration. This was true for both θ estimates of the overall full-scale score and for content area only scores. If the MBE were to use IRT scoring, it would not matter in terms of θ estimates which type of calibration was used, nor which θ estimation method was used. In terms of item fit, FS calibration required only one more item deletion than CA calibration. The mean SEMs were similar for the CA and FS calibrations, suggesting that one calibration approach does not provide more measurement precision than the other. Person fit was slightly higher for FS calibration with content area scores, but slightly lower for full-scale θ estimates. While counterintuitive, this supports the conclusion that neither FS nor CA calibration is superior to the other. However, FS calibration is easier to implement, so it might be more applicable simply for that reason.

Full-scale scoring, rather than content area scoring, was supported by the mean SEMs of the content area θ estimates and the low classical reliabilities of the content area scores. It is noteworthy that the content area SEMs were, on average, more than half the SD of examinee θ estimates, i.e., greater than

0.5. A scale of only 30 to 34 items has very imprecise measurement with the types of items on the MBE, due to the low item discriminations. Except for MLE, the SDs of the SEMs were again small, meaning that the extent of the lower precision of measurement suggested by large SEM means was at least relatively constant across examinees.

This conclusion of unidimensionality from the calibration method comparison was also supported by the factor analysis, which can be interpreted as a reparameterization of IRT. Both the item-level and subscore-level factor analyses found a very strong unrotated first factor and (at the item level) a moderately strong rotated first factor. This is especially true in the subscore-level factor analysis, which found strong first factors even though the subscores were not highly reliable in themselves (KR-20 coefficients of 0.439 to 0.648). These low reliabilities likely contributed to the fact that the correlations between subscores were only moderate, preventing an extremely strong single factor.

Estimation method also had little effect. It did not affect the mean of the θ estimates or the SEMs, but the SDs were reduced by the introduction of a normal prior for the Bayesian methods. The Bayesian methods did have slightly higher person-fit statistics, however. The Bayesian methods also produced smaller SEMs. Again, this might be due the effect of a Bayesian normal prior—it shrinks the likelihood function toward the mean somewhat, and this in turn decreases the resulting SEM, which corresponds to the SD of the likelihood function.

IRT does provide more possible scores than the 201 scores possible with NC scoring, but the correlation was so high with NC scores that it did not order examinees any differently, except for the greater specificity. However, IRT scoring facilitates scaling and equating of examinee scores, and the high correlation also suggests that pass/fail decisions will be consistent across scoring methods, which is greatly desirable. This is especially true since different jurisdictions use different cutscores.

The results of these analyses suggest that IRT calibration should be done with FS calibration rather than CA; the MBE is unidimensional enough that CA calibration does not produce different θ estimates, while introducing more complexity. Subscores should not be reported, as the reliability and mean observed SEMs of the independent content areas were too low. IRT θ estimation method would not matter with IRT scoring. The three methods correlated very highly, and the only difference of note was that MLE has a slightly higher SD than Bayesian methods, but this would disappear when transformed into scaled scores with a fixed SD.

References

- Assessment Systems Corporation (1997). *User's manual for XCALIBRE™ for Windows: Marginal-maximum-likelihood estimation program*. St. Paul, MN: Author.
- Assessment Systems Corporation (2006). *User's manual for SCOREALL*. St. Paul, MN: Author.
- Bejar, I.I. (1980). A procedure for investigating the unidimensionality of achievement tests based on item parameter estimates. *Journal of Educational Measurement*, 17, 283-296.
- Bejar, I.I. (1988). An approach to assessing unidimensionality revisited. *Applied Psychological Measurement*, 12, 377-379.
- Bejar, I.I., Weiss, D.J., & Kingsbury, G.G. (1977). *Calibration of an item pool for the adaptive measurement of achievement* (Research Report 77-5). Minneapolis, MN; University of Minnesota, Department of Psychology, Psychometric Methods Program.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gibbons, R.D., & Hedeker, D.R. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57, 423-426.
- Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory*. Boston: Kluwer.
- Reise, S. P. (1990). A comparison of item- and person-fit methods of assessing model-data fit in IRT. *Applied Psychological Measurement*, 14, 127-137.
- Segall, D. O., Moreno, K. E., & Hetter, D. H. (1997). Item pool development and evaluation. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 117-130). Washington DC: American Psychological Association.
- Thisen, D., & Wainer, H. (2001). *Test scoring*. Mahwah, NJ: Erlbaum.
- Waller, N. G. (2004). *MicroFACT 2.0 for Windows: Factor analysis for dichotomous and ordered polytomous response data*. St. Paul, MN: Assessment Systems Corporation.

Appendix A: Item Level Factor Analysis Results

Table A1: Unrotated Factor Loadings and Estimated Communalities (Squared Multiple Correlations)

Item	Factor						Estimated Communalities
	1	2	3	4	5	6	
1	0.30	-0.10	-0.09	-0.02	-0.01	-0.01	0.11
2	0.41	-0.16	0.05	0.00	-0.01	0.08	0.20
3	0.37	-0.11	-0.06	-0.01	0.03	0.06	0.16
4	0.39	-0.29	0.06	-0.03	0.00	0.16	0.27
5	0.32	0.08	-0.01	-0.02	0.00	0.01	0.11
6	0.12	-0.05	-0.03	-0.01	-0.02	-0.01	0.02
7	0.19	0.02	-0.01	0.07	0.01	-0.02	0.04
8	0.32	-0.09	-0.02	0.10	0.00	-0.10	0.13
9	0.20	0.05	0.14	-0.03	0.00	0.01	0.06
10	0.42	-0.12	0.09	-0.08	-0.01	-0.05	0.20
11	0.09	-0.25	-0.01	-0.10	0.04	-0.05	0.09
12	0.52	0.01	-0.11	-0.13	0.01	0.00	0.30
13	0.27	0.18	0.04	0.08	0.01	0.12	0.13
14	0.19	-0.05	0.10	-0.01	-0.01	0.00	0.05
15	0.14	-0.10	0.03	-0.10	-0.02	-0.06	0.05
16	0.30	-0.07	-0.06	-0.06	-0.01	0.11	0.12
17	0.21	-0.13	0.09	0.04	0.03	-0.05	0.08
18	0.33	-0.03	0.02	-0.03	0.02	-0.02	0.11
19	0.26	0.13	0.00	0.06	-0.01	0.05	0.09
20	0.55	-0.08	-0.02	-0.07	0.03	0.07	0.32
21	0.31	-0.19	0.10	0.05	-0.01	0.04	0.15
22	0.17	-0.06	0.06	-0.02	0.00	-0.11	0.05
23	0.41	-0.14	0.14	-0.09	0.01	-0.05	0.21
24	0.22	-0.11	0.09	-0.05	-0.01	-0.02	0.07
25	0.02	-0.12	0.16	0.04	0.77	0.05	0.64
26	0.43	-0.17	-0.22	-0.13	0.00	0.01	0.28
27	0.27	0.01	0.09	0.00	-0.01	0.04	0.08
28	0.42	-0.02	0.01	-0.03	0.02	0.06	0.18
29	0.32	-0.06	0.11	-0.10	0.01	0.05	0.13
30	0.27	0.03	0.10	-0.14	-0.02	0.01	0.11
31	0.12	0.02	0.02	0.07	-0.02	-0.03	0.02
32	0.44	0.19	-0.06	-0.12	0.02	-0.02	0.25
33	0.37	0.06	0.19	-0.10	-0.03	-0.01	0.19
34	0.24	0.10	0.06	0.15	-0.03	-0.05	0.10
35	0.43	0.10	0.03	-0.02	0.00	0.01	0.19
36	0.20	-0.04	-0.01	-0.05	0.00	-0.03	0.05
37	0.08	-0.06	0.09	0.03	-0.01	-0.08	0.02
38	-0.02	-0.21	0.04	0.13	0.01	-0.03	0.06
39	0.45	0.15	-0.12	-0.08	0.00	0.04	0.25
40	0.49	0.00	-0.05	0.04	0.01	-0.09	0.26
41	0.51	-0.15	-0.09	0.10	-0.01	-0.03	0.30

42	0.11	0.11	-0.06	0.18	0.01	-0.13	0.08
43	0.34	-0.09	0.01	-0.05	-0.02	0.00	0.12
44	0.39	-0.07	-0.03	-0.01	0.00	0.07	0.16
45	0.27	-0.08	0.06	0.01	-0.04	-0.19	0.12
46	0.35	0.04	0.13	0.03	-0.01	-0.01	0.14
47	0.34	0.05	0.03	0.07	0.01	-0.10	0.13
48	0.22	-0.01	0.04	-0.09	-0.01	-0.05	0.06
49	0.47	-0.06	0.27	-0.08	-0.02	-0.04	0.30
50	0.37	-0.05	-0.04	-0.02	0.00	-0.04	0.14
51	0.01	0.02	-0.10	0.06	0.00	0.06	0.02
52	0.14	-0.11	-0.10	0.00	-0.02	-0.02	0.04
53	0.46	-0.03	-0.10	0.00	0.03	0.04	0.22
54	0.44	0.01	-0.13	0.05	0.01	0.06	0.21
55	0.11	-0.08	-0.07	-0.06	0.01	0.08	0.03
56	0.27	0.00	0.01	0.07	-0.01	0.02	0.08
57	0.37	-0.10	-0.03	0.07	0.01	0.06	0.16
58	0.26	-0.07	0.13	-0.03	0.00	-0.13	0.10
59	0.17	-0.09	-0.06	0.15	-0.02	0.07	0.07
60	0.14	0.01	-0.01	0.08	0.00	-0.04	0.03
61	0.25	-0.09	-0.09	0.18	0.05	0.10	0.12
62	0.24	-0.14	-0.05	0.03	-0.02	-0.14	0.10
63	0.21	-0.05	-0.13	0.04	0.05	-0.03	0.07
64	0.24	-0.19	0.01	0.11	0.02	-0.04	0.11
65	0.45	-0.14	0.01	0.00	0.00	0.07	0.23
66	0.29	0.12	0.03	0.05	-0.02	-0.08	0.11
67	0.22	0.06	0.05	-0.11	-0.02	-0.01	0.07
68	0.20	-0.03	-0.05	0.01	0.01	-0.04	0.05
69	0.31	0.13	-0.10	-0.07	0.01	-0.07	0.13
70	0.26	0.45	0.10	0.34	0.02	0.09	0.40
71	0.34	0.07	-0.02	0.04	-0.01	0.08	0.13
72	0.16	0.03	-0.05	0.08	0.02	-0.07	0.04
73	0.29	-0.15	0.00	0.10	0.00	0.12	0.13
74	0.23	-0.01	0.15	-0.13	-0.01	0.02	0.09
75	0.10	-0.15	-0.01	-0.01	-0.01	0.12	0.05
76	0.23	0.13	-0.04	0.03	0.04	-0.09	0.08
77	0.39	0.21	-0.05	-0.04	0.02	0.01	0.20
78	0.07	-0.05	0.17	0.03	-0.02	-0.05	0.04
79	0.27	0.13	0.07	0.01	-0.01	-0.03	0.10
80	0.28	-0.01	0.10	0.03	-0.01	0.01	0.09
81	0.29	-0.08	-0.10	-0.08	0.01	0.05	0.11
82	0.26	0.14	0.02	0.08	0.03	-0.04	0.10
83	0.20	0.24	0.02	0.09	0.01	-0.02	0.11
84	0.29	0.54	0.17	0.32	0.01	0.07	0.51
85	0.23	0.00	-0.07	0.06	0.02	-0.10	0.07
86	0.33	0.23	-0.02	0.26	-0.02	-0.10	0.24
87	0.33	-0.02	0.02	-0.04	0.02	-0.01	0.11
88	0.31	-0.03	0.11	0.05	-0.02	0.11	0.12
89	0.12	-0.13	-0.04	0.07	-0.03	-0.10	0.05

90	0.40	-0.03	0.00	0.06	0.01	-0.02	0.16
91	0.23	-0.05	0.03	-0.08	0.01	-0.01	0.06
92	0.13	-0.01	0.07	0.05	-0.01	0.02	0.03
93	0.40	-0.03	-0.04	0.04	0.01	0.08	0.18
94	0.33	0.14	0.00	0.00	-0.01	0.08	0.13
95	0.38	-0.09	0.02	0.09	-0.04	0.09	0.17
96	0.15	0.04	-0.15	-0.05	0.01	0.11	0.06
97	0.25	-0.03	0.02	0.00	-0.01	-0.01	0.06
98	0.02	-0.07	-0.15	0.18	0.02	0.12	0.07
99	0.36	0.10	0.10	-0.03	0.03	0.01	0.15
100	0.41	-0.09	0.08	0.08	0.01	-0.06	0.20
101	0.16	0.15	-0.09	0.01	-0.03	-0.01	0.06
102	0.45	0.03	-0.09	-0.02	0.01	-0.05	0.21
103	0.44	-0.29	-0.05	0.09	0.01	-0.08	0.29
104	0.15	-0.02	-0.05	-0.04	-0.01	0.06	0.03
105	0.13	-0.02	-0.12	-0.04	-0.01	-0.18	0.06
106	0.21	-0.15	-0.08	0.02	0.01	0.19	0.11
107	0.13	-0.16	0.04	0.03	0.02	0.21	0.09
108	0.45	0.24	-0.08	-0.03	0.01	-0.16	0.28
109	0.33	-0.02	-0.06	0.06	-0.02	-0.07	0.12
110	0.45	-0.05	-0.20	0.05	0.00	-0.07	0.25
111	0.37	0.12	0.03	-0.10	-0.01	-0.02	0.16
112	0.28	-0.25	0.03	0.07	0.01	-0.17	0.17
113	0.01	0.02	-0.12	0.11	0.01	0.02	0.03
114	0.40	-0.15	-0.08	0.10	-0.01	-0.10	0.21
115	0.50	0.08	-0.10	-0.17	0.01	0.04	0.30
116	0.23	0.04	0.02	-0.01	0.00	0.06	0.06
117	0.27	0.21	0.01	-0.07	0.01	0.18	0.15
118	0.20	-0.05	-0.05	0.01	0.04	0.01	0.05
119	0.15	-0.08	0.09	0.07	-0.01	-0.08	0.05
120	0.29	-0.20	0.06	0.05	-0.01	-0.04	0.13
121	0.38	-0.07	0.08	-0.01	-0.02	-0.05	0.16
122	0.54	0.22	-0.11	-0.11	0.02	0.02	0.36
123	0.48	0.19	0.10	-0.06	-0.01	-0.08	0.29
124	0.49	-0.07	-0.17	-0.08	0.00	0.00	0.27
125	0.12	-0.05	0.12	0.07	-0.02	0.03	0.04
126	0.27	0.01	-0.05	0.09	0.02	-0.11	0.10
127	0.35	-0.02	-0.01	0.00	0.01	0.12	0.14
128	0.43	0.15	-0.16	-0.04	0.00	0.03	0.24
129	0.25	0.08	-0.02	0.01	0.02	0.03	0.07
130	0.28	0.01	-0.01	-0.08	0.01	-0.06	0.08
131	0.38	0.23	0.03	0.13	0.01	-0.05	0.21
132	0.27	-0.06	0.09	0.00	0.02	0.14	0.11
133	0.27	0.31	0.12	0.10	-0.04	0.00	0.19
134	0.00	0.01	-0.07	-0.03	0.01	-0.14	0.02
135	0.44	-0.01	-0.05	-0.06	0.01	-0.06	0.20
136	0.28	-0.17	-0.07	0.07	-0.01	-0.08	0.12
137	0.12	-0.02	0.04	0.09	-0.02	0.09	0.03

138	0.27	0.07	0.01	-0.11	-0.01	-0.02	0.09
139	0.33	0.04	-0.05	-0.30	-0.02	0.08	0.21
140	0.38	0.29	-0.08	-0.11	0.00	0.09	0.25
141	0.40	-0.03	-0.08	0.02	0.00	0.08	0.17
142	0.39	-0.12	-0.08	-0.03	-0.02	0.01	0.18
143	0.53	-0.15	0.00	0.03	0.00	0.05	0.31
144	0.45	0.19	-0.07	-0.06	-0.02	-0.11	0.26
145	0.27	-0.08	0.01	0.08	0.00	0.15	0.11
146	0.48	-0.07	-0.18	-0.02	0.02	0.01	0.27
147	0.26	0.17	-0.05	-0.11	0.01	0.12	0.12
148	0.52	0.18	0.09	-0.11	0.00	0.00	0.32
149	0.37	-0.10	-0.09	0.06	-0.02	-0.02	0.16
150	0.37	0.16	-0.16	-0.08	0.03	-0.02	0.19
151	0.25	0.12	-0.01	0.03	0.03	0.11	0.09
152	0.47	-0.06	-0.05	0.09	0.00	0.05	0.23
153	0.32	-0.02	-0.03	0.05	0.01	0.06	0.11
154	0.23	-0.06	0.07	-0.05	-0.03	-0.02	0.07
155	0.21	0.00	0.01	0.21	-0.03	0.15	0.11
156	0.29	-0.20	0.09	-0.01	0.01	-0.01	0.13
157	0.37	-0.02	0.27	-0.01	0.03	-0.02	0.21
158	0.25	0.12	0.10	0.02	-0.04	-0.04	0.09
159	0.04	-0.09	0.04	0.05	-0.73	0.09	0.55
160	0.45	0.07	0.08	-0.04	-0.03	-0.01	0.22
161	0.23	-0.05	0.06	-0.11	-0.01	-0.04	0.07
162	0.21	-0.09	0.08	0.07	0.03	-0.04	0.06
163	0.34	-0.14	0.01	-0.01	0.00	-0.01	0.13
164	0.48	-0.04	0.08	-0.02	0.02	0.04	0.24
165	0.09	-0.06	-0.09	0.14	-0.05	0.02	0.04
166	0.22	-0.16	0.14	-0.10	-0.04	-0.05	0.11
167	0.28	0.15	-0.09	-0.07	0.00	0.03	0.12
168	-0.04	0.18	0.20	-0.01	0.00	0.00	0.08
169	0.28	-0.09	-0.04	0.07	0.00	0.04	0.09
170	0.28	-0.01	0.05	0.21	0.01	0.00	0.12
171	0.08	0.02	-0.02	0.05	-0.02	-0.14	0.03
172	0.09	-0.12	0.07	0.00	0.02	0.03	0.03
173	0.27	-0.23	0.03	0.06	-0.03	0.08	0.13
174	0.19	-0.11	0.06	0.18	0.01	-0.03	0.09
175	0.00	-0.16	0.06	-0.03	0.02	-0.09	0.04
176	0.33	0.21	-0.06	0.10	0.01	0.05	0.17
177	0.15	-0.02	0.05	0.03	0.05	0.11	0.04
178	0.23	0.05	-0.07	-0.06	-0.02	-0.05	0.07
179	0.07	-0.15	-0.06	-0.01	-0.05	0.10	0.04
180	0.17	0.03	0.21	-0.12	0.02	0.05	0.09
181	0.20	-0.17	0.18	-0.02	0.00	0.02	0.10
182	0.39	0.04	0.02	-0.07	-0.01	-0.03	0.16
183	0.37	0.05	0.08	-0.09	-0.03	-0.05	0.16
184	0.24	0.11	-0.04	-0.04	0.01	-0.01	0.07
185	0.14	-0.22	0.00	0.03	0.01	-0.07	0.07

186	-0.03	-0.01	0.16	0.11	-0.03	0.04	0.04
187	0.31	-0.20	-0.09	0.17	-0.03	-0.03	0.17
188	-0.03	0.19	-0.08	0.13	0.01	-0.08	0.07
189	0.29	-0.01	0.00	0.16	0.01	-0.05	0.11
190	0.51	-0.06	-0.16	0.01	0.00	-0.05	0.29
191	0.23	0.05	0.04	-0.06	-0.01	-0.15	0.08
192	0.13	-0.15	0.08	0.09	-0.03	-0.01	0.05
193	0.30	0.04	-0.09	-0.07	0.01	-0.04	0.11
194	0.23	0.07	0.08	0.04	0.01	0.02	0.06
195	0.09	-0.01	-0.01	0.12	-0.04	-0.09	0.03
196	0.16	0.29	-0.15	-0.07	-0.01	-0.04	0.14
197	0.29	0.13	0.13	-0.14	-0.01	0.03	0.14
198	0.39	0.16	0.36	-0.13	-0.07	-0.03	0.33
199	0.13	-0.03	0.35	0.03	-0.05	0.04	0.14
200	0.28	-0.03	0.03	0.06	0.03	-0.01	0.09

Table A2: Varimax Rotated Factor Loadings

Item	Factor					
	1	2	3	4	5	6
1	0.20	0.26	0.04	-0.01	0.03	-0.01
2	0.21	0.31	0.17	0.01	0.17	0.00
3	0.24	0.28	0.06	-0.01	0.12	0.03
4	0.16	0.37	0.18	-0.10	0.25	0.01
5	0.28	0.12	0.08	0.09	0.04	0.00
6	0.07	0.11	0.03	-0.01	0.01	-0.02
7	0.12	0.12	0.03	0.11	0.03	0.01
8	0.14	0.30	0.08	0.12	-0.01	0.01
9	0.14	0.04	0.18	0.07	0.06	0.01
10	0.24	0.29	0.25	0.00	0.03	0.00
11	-0.02	0.22	0.08	-0.18	-0.03	0.04
12	0.46	0.28	0.09	0.00	0.03	0.01
13	0.25	0.01	0.05	0.19	0.16	0.01
14	0.09	0.12	0.16	0.03	0.05	0.00
15	0.08	0.13	0.12	-0.08	-0.04	-0.02
16	0.24	0.19	0.05	-0.05	0.14	-0.02
17	0.05	0.21	0.16	0.05	0.03	0.04
18	0.23	0.19	0.13	0.04	0.04	0.02
19	0.23	0.06	0.04	0.16	0.09	-0.01
20	0.40	0.33	0.17	0.01	0.14	0.03
21	0.09	0.29	0.18	0.03	0.14	0.00
22	0.07	0.15	0.14	0.03	-0.06	0.01
23	0.21	0.28	0.29	-0.01	0.05	0.02
24	0.09	0.17	0.17	-0.02	0.04	0.00
25	-0.08	0.05	0.09	-0.02	0.12	0.78
26	0.34	0.38	-0.01	-0.14	0.03	-0.01
27	0.18	0.11	0.15	0.07	0.10	-0.01
28	0.31	0.23	0.13	0.05	0.12	0.02

29	0.22	0.16	0.22	-0.03	0.11	0.02
30	0.24	0.06	0.21	-0.02	0.03	-0.01
31	0.06	0.08	0.04	0.10	0.01	-0.02
32	0.47	0.10	0.10	0.08	-0.02	0.02
33	0.28	0.09	0.30	0.07	0.06	-0.02
34	0.14	0.10	0.09	0.25	0.03	-0.03
35	0.36	0.15	0.14	0.13	0.07	0.00
36	0.14	0.14	0.08	-0.01	0.00	0.01
37	-0.02	0.09	0.12	0.05	-0.03	0.01
38	-0.18	0.17	0.03	0.01	0.04	0.02
39	0.47	0.15	0.03	0.08	0.04	-0.01
40	0.35	0.32	0.11	0.15	-0.01	0.01
41	0.28	0.45	0.07	0.11	0.08	-0.01
42	0.07	0.08	-0.06	0.24	-0.08	0.01
43	0.21	0.24	0.14	-0.01	0.06	-0.01
44	0.27	0.25	0.09	0.03	0.13	0.00
45	0.11	0.24	0.18	0.08	-0.10	-0.02
46	0.22	0.15	0.22	0.15	0.08	0.00
47	0.22	0.19	0.12	0.19	-0.02	0.02
48	0.17	0.10	0.13	-0.01	-0.02	0.00
49	0.26	0.23	0.42	0.07	0.08	0.00
50	0.25	0.25	0.10	0.04	0.01	0.00
51	0.02	0.02	-0.12	0.03	0.05	0.00
52	0.07	0.19	-0.03	-0.04	0.00	-0.02
53	0.35	0.29	0.05	0.05	0.10	0.03
54	0.34	0.27	-0.01	0.10	0.11	0.01
55	0.09	0.10	-0.02	-0.09	0.08	0.00
56	0.17	0.16	0.06	0.11	0.08	0.00
57	0.21	0.29	0.07	0.07	0.14	0.01
58	0.12	0.18	0.23	0.05	-0.05	0.01
59	0.04	0.19	-0.04	0.09	0.13	-0.03
60	0.07	0.11	0.01	0.11	0.00	0.00
61	0.11	0.25	-0.06	0.12	0.17	0.05
62	0.09	0.28	0.07	0.03	-0.07	-0.02
63	0.14	0.20	-0.06	0.04	0.00	0.04
64	0.03	0.30	0.08	0.06	0.05	0.02
65	0.26	0.33	0.15	0.02	0.16	0.00
66	0.23	0.11	0.10	0.18	-0.02	-0.01
67	0.21	0.04	0.14	0.00	0.00	-0.01
68	0.14	0.15	0.03	0.04	-0.01	0.01
69	0.34	0.10	0.02	0.07	-0.07	0.01
70	0.25	-0.13	0.00	0.55	0.17	0.02
71	0.27	0.14	0.05	0.12	0.13	-0.02
72	0.10	0.12	-0.01	0.12	-0.04	0.02
73	0.11	0.27	0.05	0.05	0.20	0.00
74	0.17	0.06	0.23	-0.03	0.05	0.00
75	0.02	0.14	0.03	-0.09	0.14	-0.01
76	0.22	0.07	0.03	0.15	-0.06	0.04

77	0.41	0.06	0.06	0.14	0.02	0.01
78	-0.04	0.06	0.18	0.06	0.01	0.00
79	0.24	0.05	0.13	0.15	0.02	0.00
80	0.16	0.15	0.16	0.10	0.09	0.00
81	0.24	0.21	0.02	-0.07	0.06	0.01
82	0.22	0.07	0.06	0.20	0.01	0.04
83	0.22	-0.03	0.03	0.24	0.01	0.01
84	0.31	-0.19	0.07	0.59	0.15	0.02
85	0.16	0.18	0.01	0.10	-0.06	0.02
86	0.24	0.12	0.01	0.41	-0.01	-0.02
87	0.24	0.18	0.13	0.04	0.04	0.03
88	0.17	0.16	0.16	0.09	0.19	-0.01
89	-0.01	0.21	0.01	0.04	-0.05	-0.02
90	0.25	0.26	0.10	0.13	0.07	0.01
91	0.17	0.14	0.13	-0.03	0.02	0.01
92	0.05	0.07	0.09	0.07	0.07	-0.01
93	0.28	0.25	0.06	0.08	0.15	0.01
94	0.31	0.07	0.07	0.12	0.11	-0.01
95	0.20	0.28	0.11	0.10	0.19	-0.04
96	0.20	0.06	-0.10	-0.04	0.08	0.00
97	0.16	0.16	0.10	0.05	0.05	-0.01
98	-0.03	0.12	-0.19	0.06	0.14	0.01
99	0.30	0.10	0.19	0.13	0.06	0.04
100	0.20	0.31	0.19	0.14	0.06	0.03
101	0.21	0.01	-0.05	0.10	-0.02	-0.03
102	0.37	0.26	0.06	0.09	0.00	0.00
103	0.15	0.50	0.12	0.04	0.05	0.01
104	0.14	0.09	0.00	-0.03	0.07	-0.02
105	0.11	0.14	-0.02	0.00	-0.17	-0.01
106	0.12	0.20	-0.03	-0.07	0.21	0.00
107	0.02	0.14	0.05	-0.06	0.25	0.02
108	0.45	0.13	0.07	0.21	-0.13	0.01
109	0.22	0.25	0.05	0.11	-0.01	-0.02
110	0.33	0.37	-0.04	0.09	-0.01	-0.01
111	0.35	0.09	0.16	0.07	0.01	-0.01
112	0.02	0.38	0.15	0.04	-0.05	0.03
113	0.01	0.05	-0.14	0.07	0.02	0.00
114	0.19	0.40	0.06	0.11	0.00	-0.01
115	0.50	0.19	0.09	-0.02	0.04	0.00
116	0.19	0.08	0.07	0.05	0.09	0.00
117	0.34	-0.06	0.05	0.07	0.16	0.00
118	0.14	0.16	0.01	0.02	0.04	0.03
119	0.01	0.15	0.13	0.09	0.00	0.00
120	0.07	0.31	0.16	0.03	0.06	0.01
121	0.22	0.25	0.21	0.08	0.04	-0.01
122	0.57	0.13	0.06	0.11	0.03	0.01
123	0.42	0.11	0.24	0.19	-0.02	0.00
124	0.40	0.34	0.03	-0.02	0.03	-0.01

125	0.00	0.08	0.12	0.08	0.09	-0.01
126	0.17	0.21	0.03	0.15	-0.05	0.02
127	0.26	0.19	0.08	0.04	0.17	0.01
128	0.45	0.15	-0.03	0.10	0.04	-0.01
129	0.23	0.08	0.04	0.09	0.06	0.02
130	0.23	0.14	0.10	0.01	-0.03	0.01
131	0.32	0.08	0.09	0.31	0.03	0.01
132	0.16	0.14	0.14	0.01	0.20	0.02
133	0.27	-0.08	0.13	0.31	0.05	-0.03
134	0.02	0.03	-0.03	0.00	-0.15	0.01
135	0.35	0.26	0.12	0.05	-0.01	0.01
136	0.10	0.33	0.04	0.04	0.00	-0.01
137	0.04	0.07	0.03	0.08	0.13	-0.02
138	0.26	0.07	0.12	0.01	-0.01	0.00
139	0.39	0.07	0.12	-0.17	0.04	-0.02
140	0.49	-0.02	0.02	0.10	0.06	-0.01
141	0.29	0.25	0.03	0.06	0.13	-0.01
142	0.26	0.32	0.07	-0.02	0.06	-0.02
143	0.31	0.40	0.16	0.05	0.16	0.00
144	0.45	0.14	0.09	0.15	-0.09	-0.02
145	0.14	0.19	0.05	0.05	0.22	0.00
146	0.37	0.36	-0.01	0.02	0.06	0.01
147	0.34	-0.02	0.02	0.02	0.09	0.00
148	0.48	0.11	0.24	0.13	0.05	0.00
149	0.22	0.32	0.03	0.07	0.06	-0.02
150	0.41	0.12	-0.02	0.07	-0.03	0.02
151	0.25	0.04	0.02	0.11	0.13	0.02
152	0.30	0.33	0.07	0.13	0.14	0.00
153	0.22	0.20	0.05	0.08	0.12	0.01
154	0.13	0.14	0.16	0.00	0.03	-0.02
155	0.08	0.13	-0.01	0.18	0.22	-0.03
156	0.09	0.28	0.19	-0.03	0.08	0.02
157	0.19	0.16	0.36	0.12	0.10	0.05
158	0.20	0.04	0.15	0.16	0.02	-0.03
159	-0.06	0.08	0.07	0.01	0.12	-0.72
160	0.35	0.17	0.21	0.12	0.06	-0.02
161	0.16	0.13	0.17	-0.04	-0.01	0.00
162	0.05	0.18	0.13	0.09	0.04	0.04
163	0.18	0.28	0.13	0.01	0.06	0.01
164	0.32	0.26	0.21	0.07	0.13	0.03
165	0.01	0.14	-0.08	0.08	0.06	-0.06
166	0.08	0.18	0.25	-0.07	0.02	-0.03
167	0.33	0.05	0.00	0.06	0.01	-0.01
168	0.01	-0.20	0.15	0.12	0.01	0.02
169	0.15	0.24	0.03	0.06	0.11	0.00
170	0.10	0.21	0.07	0.23	0.11	0.02
171	0.04	0.08	0.02	0.09	-0.11	-0.01
172	-0.01	0.11	0.10	-0.03	0.07	0.03

173	0.06	0.30	0.11	-0.02	0.17	-0.03
174	-0.01	0.23	0.08	0.15	0.08	0.02
175	-0.09	0.11	0.10	-0.07	-0.06	0.03
176	0.33	0.07	-0.01	0.23	0.08	0.01
177	0.09	0.07	0.06	0.03	0.15	0.05
178	0.23	0.11	0.03	0.03	-0.05	-0.03
179	0.01	0.13	-0.02	-0.10	0.11	-0.06
180	0.13	-0.02	0.26	-0.02	0.08	0.03
181	0.02	0.18	0.24	-0.02	0.10	0.01
182	0.32	0.17	0.16	0.06	0.02	0.00
183	0.30	0.14	0.21	0.06	0.00	-0.02
184	0.25	0.06	0.04	0.07	0.00	0.01
185	-0.02	0.26	0.07	-0.03	-0.01	0.01
186	-0.12	-0.03	0.10	0.10	0.09	-0.02
187	0.08	0.38	0.00	0.09	0.08	-0.03
188	0.03	-0.07	-0.12	0.19	-0.09	0.01
189	0.13	0.22	0.05	0.19	0.04	0.02
190	0.38	0.38	0.03	0.07	0.02	0.00
191	0.18	0.10	0.14	0.06	-0.12	0.00
192	-0.03	0.19	0.11	0.05	0.07	-0.02
193	0.29	0.15	0.03	0.02	-0.03	0.00
194	0.16	0.07	0.11	0.12	0.07	0.02
195	0.00	0.11	0.00	0.13	-0.04	-0.03
196	0.32	-0.09	-0.09	0.10	-0.10	-0.02
197	0.29	-0.01	0.22	0.04	0.05	-0.01
198	0.30	-0.01	0.46	0.14	0.05	-0.04
199	-0.03	0.01	0.34	0.10	0.14	-0.02
200	0.16	0.19	0.09	0.11	0.06	0.03

**Appendix B: MBE FS and CA
IRT Item Parameter Estimates**

Item	<i>a</i>		<i>b</i>		<i>c</i>	
	FS	CA	FS	CA	FS	CA
1	0.78	0.36	0.19	-1.52	0.63	0.35
2	0.47	0.47	-1.07	-1.12	0.21	0.22
3	0.60	0.48	0.14	-0.14	0.35	0.28
4	0.50	0.49	-0.45	-0.57	0.26	0.23
5	0.36	0.32	-2.08	-2.29	0.24	0.26
6	0.37	0.39	3.00	3.00	0.27	0.28
7	0.32	0.26	1.96	2.52	0.26	0.27
8	0.47	0.36	0.64	0.73	0.26	0.25
9	0.30		-3.00		0.07	
10	0.51	0.47	-2.05	-2.28	0.28	0.27
11	0.20		2.56		0.36	
12	0.74	0.68	-0.27	-0.36	0.21	0.20
13	0.33	0.30	0.26	0.31	0.22	0.23
14	0.32	0.30	1.22	1.16	0.34	0.32
15	0.20	0.20	0.53	0.63	0.28	0.29
16	0.49	0.45	1.13	1.16	0.23	0.21
17	0.24	0.26	-0.77	-0.74	0.25	0.25
18	0.37	0.33	-1.84	-2.06	0.24	0.26
19	0.29	0.23	-1.57	-1.86	0.25	0.27
20	0.78	0.74	-0.68	-0.81	0.23	0.21
21	0.35	0.31	-0.79	-0.86	0.22	0.23
22	0.20	0.20	-2.61	-2.86	0.25	0.23
23						
24	0.24	0.24	-1.56	-1.51	0.25	0.26
25	0.73	0.63	3.00	3.00	0.53	0.52
26	0.66	0.58	1.12	1.29	0.12	0.12
27	0.30	0.27	-0.78	-0.71	0.21	0.24
28	0.51	0.48	-1.19	-1.29	0.23	0.23
29	0.35	0.33	-2.36	-2.45	0.21	0.23
30	0.29	0.28	-2.41	-2.45	0.21	0.25
31						
32	0.53	0.43	-1.48	-1.78	0.23	0.24
33	0.42	0.36	-2.30	-2.61	0.22	0.24
34	0.27	0.23	-1.49	-2.36	0.25	0.13
35	0.49	0.42	-1.92	-2.43	0.21	0.12
36	0.23	0.23	-2.86	-2.99	0.26	0.14
37		0.20		-2.99		0.12
38	1.26	1.21	2.79	3.00	0.30	0.30
39	0.56	0.42	-0.71	-1.16	0.20	0.12
40	0.64	0.57	-0.93	-1.29	0.23	0.13
41	0.74	0.62	0.82	0.86	0.10	0.07
42	0.50		2.75		0.36	
43	0.39	0.31	-0.93	-1.54	0.23	0.13

44	0.48	0.38	0.37	0.28	0.16	0.12
45	0.29	0.24	-1.27	-2.02	0.23	0.13
46		0.43		-2.99		0.05
47	0.39	0.31	-0.88	-1.51	0.23	0.13
48	0.23	0.20	-1.73	-2.73	0.23	0.12
49	0.57		-2.90		0.23	
50	0.54	0.40	1.03	1.07	0.17	0.12
51	1.23	1.53	3.00	3.00	0.08	0.07
52	0.43		2.47		0.35	
53	0.59	0.47	0.00	-0.19	0.16	0.11
54	0.68	0.49	0.56	0.46	0.20	0.14
55	0.62	0.61	3.00	3.00	0.20	0.20
56	0.33	0.26	0.83	0.59	0.18	0.11
57	0.45	0.37	-0.71	-1.28	0.24	0.13
58						
59	0.52	0.22	2.09	1.75	0.35	0.16
60	0.33		2.14		0.38	
61	0.59	0.42	2.18	2.54	0.16	0.12
62	0.38	0.24	0.90	0.01	0.31	0.14
63	0.59	0.40	2.60	3.00	0.15	0.12
64	0.42	0.24	0.99	-0.12	0.36	0.14
65	0.56	0.46	-0.57	-0.93	0.21	0.13
66	0.33	0.27	-1.51	-2.31	0.26	0.13
67	0.23	0.20	-1.56	-2.35	0.21	0.12
68	0.27	0.20	0.15	-1.70	0.30	0.07
69	0.36	0.31	-2.36	-2.99	0.25	0.05
70	0.28	0.48	-0.20	-0.44	0.18	0.08
71	0.39	0.31	-1.06	-1.81	0.22	0.09
72	0.41	0.32	3.00	3.00	0.16	0.12
73	0.42	0.22	1.45	1.85	0.17	0.10
74	0.31	0.20	-3.00	-2.95	0.05	0.40
75	0.48	0.74	3.00	3.00	0.26	0.29
76	0.32	0.24	1.09	0.45	0.22	0.09
77	0.44	0.36	-1.51	-2.19	0.21	0.09
78		0.33		1.82		0.79
79	0.31	0.33	-2.89	-2.99	0.24	0.05
80	0.31	0.23	-0.18	-0.74	0.18	0.09
81	0.40	0.20	0.42	-0.57	0.27	0.10
82	0.30	0.31	-2.74	-2.99	0.25	0.05
83	0.25	0.25	1.33	0.51	0.20	0.09
84	0.31	0.56	-1.27	-1.05	0.19	0.10
85	0.40	0.25	1.78	1.52	0.22	0.10
86	0.42	0.44	-0.35	-0.90	0.27	0.11
87	0.38	0.29	-1.99	-2.99	0.25	0.05
88	0.34	0.31	-2.45	-2.99	0.23	0.05
89		0.22		-2.35		0.06
90	0.50	0.32	-0.50	-1.34	0.25	0.09
91	0.26	0.20	-1.49	-2.99	0.23	0.08

92		0.20		-2.99		0.05
93	0.57	0.37	-0.30	-1.18	0.30	0.11
94	0.39	0.34	-0.71	-1.33	0.24	0.10
95	0.45	0.32	-1.96	-2.99	0.26	0.05
96	0.67	0.77	3.00	3.00	0.10	0.11
97	0.30	0.20	0.30	-0.52	0.22	0.09
98	1.21	1.52	3.00	3.00	0.05	0.05
99	0.42	0.42	-2.88	-2.99	0.22	0.05
100	0.53	0.33	-1.04	-2.22	0.30	0.10
101	0.20	0.20	0.47	-0.64	0.27	0.13
102	0.57	0.50	-0.33	-0.62	0.20	0.12
103	0.75	0.49	0.17	-0.38	0.32	0.16
104	0.44	0.20	1.47	-2.99	0.54	0.05
105	0.61		2.04		0.47	
106	0.77	0.66	2.57	3.00	0.12	0.12
107	0.68	0.58	2.12	2.65	0.42	0.43
108	0.54	0.51	-1.21	-1.55	0.22	0.12
109	0.47	0.34	0.27	-0.31	0.27	0.13
110	0.67	0.57	1.07	1.12	0.10	0.08
111	0.43	0.37	-2.41	-2.99	0.23	0.13
112	0.31	0.27	-2.13	-2.88	0.26	0.13
113	1.01	1.05	3.00	3.00	0.18	0.17
114	0.63	0.44	0.10	-0.48	0.31	0.16
115	0.61	0.55	-0.51	-0.69	0.15	0.11
116	0.27	0.22	-0.66	-1.37	0.23	0.13
117	0.30	0.25	0.01	-0.35	0.19	0.12
118	0.38	0.24	1.51	0.62	0.32	0.14
119	0.20	0.20	-3.00	-2.99	0.15	0.12
120	0.34	0.28	-0.52	-1.14	0.24	0.13
121	0.44	0.38	-2.44	-2.99	0.24	0.13
122	0.65	0.65	-0.94	-1.06	0.14	0.10
123						
124	0.66	0.56	0.53	0.54	0.12	0.10
125						
126	0.47	0.27	0.33	-1.26	0.43	0.14
127	0.43	0.32	0.00	-0.38	0.20	0.12
128	0.65	0.51	0.12	-0.25	0.27	0.16
129	0.34	0.25	0.67	0.11	0.25	0.13
130	0.30	0.26	-2.06	-2.79	0.23	0.13
131	0.42	0.38	-1.14	-1.47	0.18	0.12
132	0.33	0.24	-0.19	-0.88	0.24	0.14
133	0.28	0.24	-1.36	-1.90	0.19	0.12
134	0.20	0.20	3.00	3.00	0.39	0.37
135	0.54	0.47	-1.06	-1.44	0.24	0.15
136	0.50	0.28	0.45	-0.71	0.41	0.20
137		0.21		2.53		0.39
138	0.29	0.26	-1.83	-2.29	0.22	0.16
139	0.37	0.33	-2.47	-2.94	0.22	0.16

140	0.41	0.39	-0.69	-0.82	0.15	0.13
141	0.61	0.50	1.52	1.75	0.10	0.09
142	0.58	0.53	1.10	1.15	0.14	0.13
143	0.20	0.67	3.00	0.00	0.43	0.12
144	0.52	0.46	-1.19	-1.43	0.17	0.13
145	0.47	0.39	1.77	1.88	0.19	0.16
146	0.78	0.69	1.95	2.21	0.05	0.05
147	0.30	0.26	-0.61	-1.11	0.24	0.16
148	0.63	0.60	-1.04	-1.24	0.18	0.13
149	0.71	0.62	1.30	1.48	0.17	0.17
150	0.45	0.40	0.51	0.45	0.15	0.12
151	0.31	0.26	0.47	0.24	0.21	0.16
152	0.62	0.52	0.88	1.02	0.09	0.09
153	0.42	0.36	0.77	0.70	0.18	0.14
154	0.25	0.24	0.09	-0.23	0.19	0.14
155	0.42	0.30	2.52	3.00	0.19	0.17
156	0.44	0.34	1.00	0.63	0.25	0.16
157	0.40	0.42	-2.36	-2.43	0.19	0.14
158	0.27	0.25	0.82	0.69	0.15	0.12
159	0.52	0.50	3.00	3.00	0.51	0.50
160	0.54	0.53	-1.15	-1.37	0.21	0.15
161	0.24	0.22	-0.14	-0.41	0.18	0.14
162	0.24	0.20	-1.38	-2.24	0.27	0.17
163	0.47	0.42	0.38	0.18	0.26	0.21
164	0.58	0.49	-1.54	-1.94	0.22	0.16
165	0.71	0.70	2.83	3.00	0.26	0.27
166	0.24	0.23	-1.83	-2.37	0.25	0.17
167	0.35	0.26	-0.18	0.28	0.25	0.33
168	0.22	0.21	3.00	3.00	0.75	0.74
169	0.38	0.34	1.28	1.86	0.18	0.22
170	0.34	0.32	-0.04	0.33	0.23	0.29
171	0.56	0.68	3.00	3.00	0.27	0.27
172	0.31	0.40	3.00	3.00	0.38	0.40
173	0.47	0.32	1.03	1.20	0.32	0.29
174	0.25	0.27	0.21	0.39	0.29	0.31
175		0.20		-2.99		0.31
176	0.49	0.33	0.43	0.86	0.27	0.31
177	0.25	0.27	2.58	3.00	0.25	0.31
178	0.28	0.27	0.37	1.28	0.21	0.32
179	0.58	0.71	3.00	3.00	0.28	0.29
180		0.20		-2.99		0.43
181	0.21	0.20	-2.17	-2.99	0.25	0.20
182	0.48	0.36	-0.29	-0.20	0.22	0.26
183	0.44	0.38	-1.89	-2.04	0.26	0.32
184	0.28	0.22	-1.48	-1.13	0.26	0.37
185	0.20	0.21	0.41	1.21	0.29	0.37
186	0.20	0.21	3.00	3.00	0.70	0.70
187	0.64	0.48	1.46	1.67	0.22	0.19

188	1.14	1.07	3.00	3.00	0.30	0.29
189	0.48	0.33	0.22	-0.18	0.41	0.34
190	0.72	0.65	-0.03	0.12	0.18	0.23
191	0.37		-3.00		0.08	
192	0.20	0.20	-3.00	-2.99	0.23	0.27
193	0.34	0.27	-0.47	-0.03	0.20	0.29
194	0.25	0.25	-2.71	-2.99	0.24	0.05
195	0.48	0.49	3.00	3.00	0.35	0.35
196	0.28	0.41	3.00	3.00	0.18	0.25
197	0.39		-3.00		0.05	
198		0.20		2.57		0.95
199		0.20		-2.99		0.61
200	0.33	0.30	0.01	0.38	0.21	0.27

**Appendix C: Descriptive Statistics of Item Parameters
Within Each Content Area (Excluding Deleted Items)**

Parameter	Statistic	Constitutional Law	Contracts	Criminal Law	Evidence	Real Property	Torts
<i>a</i>	Mean	0.39	0.43	0.37	0.40	0.40	0.35
	SD	0.14	0.29	0.25	0.19	0.15	0.19
	Skewness	0.90	2.69	3.47	1.45	0.41	2.18
	Kurtosis	0.26	8.06	14.45	3.06	-0.88	5.52
<i>b</i>	Mean	-0.57	-0.40	-0.91	-0.71	0.12	0.55
	SD	1.68	2.02	2.07	1.67	1.80	2.19
	Skewness	0.71	0.45	0.76	0.65	0.11	-0.47
	Kurtosis	-0.30	-1.02	-0.65	0.38	-1.07	-1.04
<i>c</i>	Mean	0.26	0.13	0.12	0.14	0.17	0.35
	SD	0.06	0.04	0.14	0.06	0.09	0.17
	Skewness	2.30	2.22	4.05	4.15	2.32	1.94
	Kurtosis	9.89	9.49	18.00	20.93	5.91	4.48

Appendix D: Correlation Matrices for θ Estimates from Each Content Area

Table D1: Correlation Matrix of θ Estimates for Constitutional Law

	MAP CA	MLE CA	EAP CA	MAP FS	MLE FS	EAP FS	NC CA
MAP CA							
MLE CA	0.982						
EAP CA	0.995	0.983					
MAP FS	0.964	0.948	0.962				
MLE FS	0.954	0.955	0.955	0.990			
EAP FS	0.957	0.947	0.963	0.994	0.990		
NC CA	0.953	0.938	0.949	0.961	0.953	0.954	

Table D2: Correlation Matrix of θ Estimates for Contracts

	MAP CA	MLE CA	EAP CA	MAP FS	MLE FS	EAP FS	NC CA
MAP CA							
MLE CA	0.988						
EAP CA	0.974	0.988					
MAP FS	0.931	0.923	0.902				
MLE FS	0.916	0.929	0.893	0.975			
EAP FS	0.920	0.922	0.920	0.977	0.973		
NC CA	0.959	0.948	0.934	0.943	0.923	0.929	

Table D3: Correlation Matrix of θ Estimates for Criminal Law

	MAP CA	MLE CA	EAP CA	MAP FS	MLE FS	EAP FS	NC CA
MAP CA							
MLE CA	0.994						
EAP CA	0.986	0.995					
MAP FS	0.972	0.966	0.959				
MLE FS	0.968	0.969	0.955	0.992			
EAP FS	0.971	0.968	0.961	0.997	0.994		
NC CA	0.979	0.976	0.966	0.965	0.963	0.966	

Table D4: Correlation Matrix of θ Estimates for Evidence

	MAP CA	MLE CA	EAP CA	MAP FS	MLE FS	EAP FS	NC CA
MAP CA							
MLE CA	0.988						
EAP CA	0.976	0.987					
MAP FS	0.979	0.968	0.954				
MLE FS	0.973	0.977	0.946	0.993			
EAP FS	0.973	0.968	0.962	0.992	0.992		
NC CA	0.969	0.958	0.943	0.962	0.957	0.956	

Table D5: Correlation Matrix of θ Estimates for Real Property

	MAP CA	MLE CA	EAP CA	MAP FS	MLE FS	EAP FS	NC CA
MAP CA							
MLE CA	0.988						
EAP CA	0.960	0.988					
MAP FS	0.988	0.976	0.949				
MLE FS	0.984	0.987	0.953	0.993			
EAP FS	0.968	0.976	0.969	0.979	0.993		
NC CA	0.969	0.956	0.928	0.971	0.965	0.950	

Table D6: Correlation Matrix of θ Estimates for Torts

	MAP CA	MLE CA	EAP CA	MAP FS	MLE FS	EAP FS	NC CA
MAP CA							
MLE CA	0.987						
EAP CA	0.987	0.987					
MAP FS	0.972	0.969	0.958				
MLE FS	0.902	0.921	0.896	0.923			
EAP FS	0.913	0.899	0.915	0.928	0.927		
NC CA	0.935	0.927	0.924	0.917	0.868	0.852	

Appendix E: Descriptive Statistics of IRT Person Estimates for Each Content Area

Table E1: Descriptive Statistics for Constitutional Law Estimates

	Calibration and Estimation	Minimum	Maximum	Mean	SD
θ Estimates	MAP, CA	-3.05	2.21	0.04	0.77
	EAP, CA	-2.97	2.07	0.03	0.79
	MLE, CA	-5.99	5.89	0.04	1.37
	MAP, FS	-3.09	2.24	0.03	0.81
	EAP, FS	-2.59	2.07	0.02	0.82
	MLE, FS	-5.97	5.77	0.07	1.40
SEM	MAP, CA	.49	.83	0.57	0.04
	EAP, CA	.51	1.86	0.59	0.05
	MLE, CA	.56	8.80	0.79	0.25
	MAP, FS	.51	.72	0.59	0.03
	EAP, FS	.52	2.31	0.60	0.04
	MLE, FS	.59	4.73	0.79	0.18
Person Fit	MAP, CA	-4.28	2.12	0.04	0.86
	EAP, CA	-4.21	2.15	0.06	0.84
	MLE, CA	-4.26	2.58	-0.04	0.74
	MAP, FS	-4.55	2.32	0.16	0.78
	EAP, FS	-3.89	2.47	0.19	0.79
	MLE, FS	-4.48	2.76	0.09	0.71

Table E2: Descriptive Statistics for Contracts Estimates

	Calibration and Estimation	Minimum	Maximum	Mean	SD
θ Estimates	MAP, CA	-2.69	3.12	-0.05	0.69
	EAP, CA	-2.38	2.91	-0.09	0.72
	MLE, CA	-6.00	5.31	-0.17	1.18
	MAP, FS	-2.67	3.14	0.05	0.82
	EAP, FS	-2.47	3.05	0.01	0.82
	MLE, FS	-6.00	5.88	0.00	1.33
SEM	MAP, CA	0.44	1.33	0.52	0.06
	EAP, CA	0.46	0.84	0.55	0.05
	MLE, CA	0.49	4.34	0.70	0.25
	MAP, FS	0.50	0.88	0.58	0.03
	EAP, FS	0.52	1.98	0.59	0.03
	MLE, FS	0.52	9.00	0.77	0.24

Person Fit	MAP, CA	-5.19	2.71	0.09	0.91
	EAP, CA	-5.03	2.73	0.11	0.91
	MLE, CA	-5.04	2.77	0.02	0.89
	MAP, FS	-5.24	2.68	0.20	0.92
	EAP, FS	-4.80	2.71	0.23	0.92
	MLE, FS	-4.70	2.85	0.15	0.89

Table E3: Descriptive Statistics for Criminal Law Estimates

	Calibration and Estimation	Minimum	Maximum	Mean	SD
θ Estimates	MAP, CA	-3.03	3.13	0.00	0.81
	EAP, CA	-2.80	2.83	0.00	0.80
	MLE, CA	-5.80	5.86	0.06	1.44
	MAP, FS	-2.97	3.05	0.02	0.79
	EAP, FS	-2.56	2.73	0.01	0.78
	MLE, FS	-5.99	5.24	0.03	1.39
SEM	MAP, CA	0.55	4.13	0.67	0.05
	EAP, CA	0.58	0.96	0.66	0.03
	MLE, CA	0.59	2.42	0.87	0.10
	MAP, FS	0.59	1.03	0.64	0.03
	EAP, FS	0.60	0.87	0.65	0.02
	MLE, FS	0.68	5.19	0.86	0.14
Person Fit	MAP, CA	-4.55	2.27	0.20	0.87
	EAP, CA	-4.56	2.28	0.20	0.87
	MLE, CA	-4.51	2.85	0.18	0.83
	MAP, FS	-4.78	2.16	0.15	0.87
	EAP, FS	-4.77	2.21	0.17	0.87
	MLE, FS	-4.68	2.83	0.14	0.81

Table E4: Descriptive Statistics for Evidence Estimates

	Calibration and Estimation	Minimum	Maximum	Mean	SD
θ Estimates	MAP, CA	-2.88	2.79	-0.04	0.82
	EAP, CA	-2.64	2.54	-0.05	0.84
	MLE, CA	-5.92	5.95	-0.07	1.36
	MAP, FS	-2.63	2.80	0.04	0.84
	EAP, FS	-2.35	2.63	0.02	0.85
	MLE, FS	-6.00	5.61	0.01	1.36

SEM	MAP, CA	0.51	0.77	0.56	0.04
	EAP, CA	0.52	2.12	0.57	0.04
	MLE, CA	0.59	4.34	0.73	0.15
	MAP, FS	0.49	0.72	0.56	0.03
	EAP, FS	0.50	2.02	0.57	0.03
	MLE, FS	0.57	8.63	0.74	0.26
Person Fit	MAP, CA	-4.29	2.65	0.15	0.87
	EAP, CA	-4.29	2.66	0.16	0.86
	MLE, CA	-4.35	2.66	0.06	0.82
	MAP, FS	-4.34	2.68	0.19	0.85
	EAP, FS	-4.29	2.72	0.23	0.85
	MLE, FS	-4.39	2.65	0.10	0.81

Table E5: Descriptive Statistics for Real Property Estimates

	Calibration and Estimation	Minimum	Maximum	Mean	SD
θ Estimates	MAP, CA	-2.94	2.94	0.03	0.88
	EAP, CA	-2.52	2.71	0.01	0.88
	MLE, CA	-5.96	5.51	0.00	1.42
	MAP, FS	-2.75	2.90	0.04	0.85
	EAP, FS	-2.38	2.72	0.02	0.84
	MLE, FS	-5.99	5.97	0.00	1.35
SEM	MAP, CA	0.54	0.70	0.58	0.02
	EAP, CA	0.54	2.38	0.59	0.03
	MLE, CA	0.63	6.15	0.76	0.18
	MAP, FS	0.51	0.75	0.57	0.03
	EAP, FS	0.52	2.34	0.58	0.04
	MLE, FS	0.58	7.61	0.75	0.27
Person Fit	MAP, CA	-4.95	2.78	0.23	0.93
	EAP, CA	-4.95	2.80	0.24	0.93
	MLE, CA	-5.02	2.90	0.12	0.90
	MAP, FS	-4.81	2.66	0.24	0.93
	EAP, FS	-4.81	2.68	0.26	0.93
	MLE, FS	-4.79	2.86	0.15	0.90

Table E6: Descriptive Statistics for Torts Estimates

	Calibration and Estimation	Minimum	Maximum	Mean	SD
θ Estimates	MAP, CA	-2.24	2.90	0.33	0.63
	EAP, CA	-1.92	2.59	0.29	0.65
	MLE, CA	-6.00	5.36	0.39	1.52
	MAP, FS	-2.64	2.76	0.05	0.72
	EAP, FS	-2.22	2.52	0.01	0.72
	MLE, FS	-5.99	5.77	0.01	1.42
SEM	MAP, CA	0.57	1.03	0.65	0.05
	EAP, CA	0.59	1.08	0.68	0.04
	MLE, CA	0.67	9.00	1.00	0.57
	MAP, FS	0.60	0.83	0.67	0.03
	EAP, FS	0.61	0.81	0.68	0.03
	MLE, FS	0.67	7.62	0.95	0.22
Person Fit	MAP, CA	-4.03	2.73	0.09	0.94
	EAP, CA	-3.96	2.78	0.15	0.92
	MLE, CA	-3.92	2.94	-0.07	0.85
	MAP, FS	-4.30	2.53	0.18	0.90
	EAP, FS	-4.10	2.58	0.21	0.90
	MLE, FS	-4.10	2.85	0.13	0.87