

Calibrating Item Parameters Using Small Samples A Comparison of Methods Using a Foreign Language Assessment

Authors

Tia M. Fechter, Ed.D.

W. Alan Nicewander, Ph.D.

Organization

Defense Testing and Assessment Center

Abstract

In response to an expressed need to reduce the number of examinees used for field testing multiple-choice items, this study compares various calibration models thought to improve accuracy of item parameter estimates for small examinee sample sizes. Calibration methods range from fully estimated three-parameter logistic models to fully restricted approaches. Results of this study revealed that using prior information about both difficulty and discrimination of items enhanced the accuracy to which item parameter estimates were recovered, but not enough to adequately recover item characteristic curves or their parameters. However, test characteristic curves were recovered with moderate accuracy under small sample size conditions. It is recommended that no fewer than 100 examinees be used for field testing multiple-choice items and constructing parallel forms. Item-specific priors should be used during item calibration. Small sample sizes should not be used for obtaining item parameter estimates for computer adaptive testing environments.

CLEARED
For Open Publication

Mar 16, 2023

Department of Defense
OFFICE OF PREPUBLICATION AND SECURITY REVIEW

Executive Summary

In response to an expressed need to reduce the number of examinees used for field testing multiple-choice items, this study compared six calibration models (labeled Condition 0 through Condition 5) thought to improve accuracy of item parameter estimates for small examinee sample sizes. Calibration methods range from fully estimated three-parameter logistic models to fully restricted approaches. Examinee sample sizes of 25, 50, 75, 100, 150, 200, and 1,000 were randomly drawn from the larger pool of 4,187 examinees, who had responded to the same set of items on a Spanish listening DLPT5 language assessment. Item parameters were estimated using the six calibration models. Analyses were replicated 100 times for each sample size condition using a different randomly drawn set of examinees. Summary statistics were aggregated across replications and served to summarize the difference between control and study condition statistics. Control statistics were obtained by calibrating item parameter estimates using the complete examinee pool. The resulting estimates were considered “trustworthy” as they were obtained using an adequate number of examinees. Analyses conducted include calibration failure rates, item parameter, item characteristic curve and test characteristic curve recovery, model fit, and classification accuracy and consistency. Results from each of these analyses are described below and recommendations for operational practice follow.

Failed Calibrations

Calibrations can fail when all examinees sampled either respond incorrectly or correctly to at least one item on the assessment. This occurrence resulted in an overwhelming number of replications failing to converge at sample sizes of 25 (88%) and 50 (43%) and a less overwhelming number for sample sizes of 75 (11%) and 100 (6%). Replications failing to converge were dropped from the study. The consequence of including too few examinees in field testing is lack of representativeness of the sample chosen to the target population.

Parameter Recovery

Of all the conditions modeled, Condition 0 item-specific priors resulted in the best fit between study and control test characteristic curves, especially in the presence of small sample sizes (200 or smaller). At the level of the item characteristic curve, Condition 0 consistently resulted in the lowest average difference across sample size conditions. However, the item parameter estimates were considered unacceptable for all sample sizes less than 1,000 under all models implemented.

Model Fit

Model fit was evaluated for each of the models using the full set of examinee data. Both models that estimate all parameters (Conditions 0 and 5) fit the data best and are seemingly indistinguishable in terms of fit. All other models fit significantly worse than Conditions 0 and 5.

Classification

Classification consistency of examinees into ILR levels may be considered the most important evaluation criterion as the decisions made based on assigned ILR levels are high stakes in nature and come with potentially high costs and drastic consequences for errors made. Differences in decision consistency were apparent such that, in general, larger sample sizes resulted in higher decision consistency, as expected. Condition 0 resulted in the highest classification consistency across all sample sizes and ranged from a high of 94% (92% when adjusted for chance agreement) for a sample size of 1,000 to a low of 70% (58%, adjusted) for a sample size of 25.

Recommendations and Cautions

To the extent that results from this study are generalizable to other language assessments, skills, test lengths, examinee ability distributions, and cut scores, the following recommendations and cautions are made for operational practice:

1. Use prior information (based on classical analyses of test data) about both difficulty (p -values) and discrimination (item-total biserial correlations corrected for guessing) of items to calibrate item parameter estimates as implemented in Condition 0.
2. It is recommended that no fewer than 100 examinees be used for field testing multiple-choice items and constructing parallel forms.
3. Do not use individual item parameter estimates obtained using less than 1,000 examinees to assess item quality. For example, it is not appropriate to use the difficulty index (i.e., b -parameter estimate) to exclude items from a test form due to the mismatch of difficulty and ILR level designation.
4. Do not use individual item parameter estimates for constructing computer adaptive test (CAT) item banks. More examinee response data would need to be collected and item recalibration would need to take place before items could be used in a CAT testing environment.

Calibrating Item Parameters Using Small Samples: A Comparison of Methods Using a Foreign Language Assessment

The DLPT5 testing system provides computer-delivered assessments designed to measure foreign language proficiency in reading and listening of native English speakers or those with strong English skills for the purpose of making training, graduation, job placement or assignment, and incentive pay decisions for civilian and military personnel with language skills in the United States government. Specifically, the lower range DLPT5 assessments are meant to classify examinees into one of seven Interagency Language Roundtable (ILR) proficiency levels (i.e., 0, 0+, 1, 1+, 2, 2+, and 3). Therefore, it is important that ability be estimated with great accuracy, especially due to the number of high stakes uses for which the assessment is designed, in addition to the number of classifications that exist. It is only with accurate and precise measurement that examinees can be adequately classified as uncertainty in item parameter estimates has been shown to underestimate the variance of ability resulting in misleading confidence intervals around the ability estimate (Tsutakawa & Johnson, 1990).

The current issue that DLPT5 faces is that there are several languages for which the population of potential examinees is smaller than the preferred sample size of 1,000 examinees for estimating item parameters using more sophisticated models like the three-parameter logistic model (3-PLM). Even for languages where larger populations exist, the cost incurred for field testing items is substantial, namely due to the need to travel the world on a number of occasions to build the examinee pool. Further, as items are field tested, items become exposed and face the danger that the DLPT5 testing system experiences a breach in security. Thus, it is desirable that fewer examinees be used for the field testing of items. However, it was shown in the review of the literature (Sukin & Nicewander, 2012) that while some promising alternatives exist to calibrating item parameter estimates with small sample sizes, namely the use of modified IRT models and Bayesian item calibration approaches, none of the methods adequately resembled the accuracy to which large sample sizes (e.g., at least 1,000 examinees) provide. Regardless, there is reason to conduct more research with the promising methods to try to remedy some of the known issues. This study aims to compare calibration methods using various sample sizes. It is anticipated that results of this work will lead to the improvement of calibration of item parameter estimates given the constraints of the DLPT5 testing system.

Methodology

This section describes the calibration methods compared along with the data and sample size conditions implemented. In addition, the evaluation criteria used for comparing methods are described.

Calibration Methods

All of the item response theory (IRT) models proposed in this study are based on the three-parameter logistic model (3-PLM), which can be defined as

$$P(u_i = 1|a_i, b_i, c_i, \theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}},$$

where $P(u_i = 1|a_i, b_i, c_i, \theta)$ is the probability of an examinee with ability equal to θ responding correctly ($u_i = 1$) to an item (i) with parameters a_i , b_i , and c_i . Parameters a_i describe the discrimination, b_i describe the difficulty, and c_i describe the lower asymptote for i . This study compares two unrestricted models (Condition 0 and Condition 5) and four modifications (Conditions 1–4) of this model for estimating item parameters (see Appendix A):

Conditions 0 and 5: unrestricted 3-PLM (3-PLM; a -, b -, and c - parameters are estimated),

Condition 1: restricted 3-PLM with each item's discrimination parameter (a) fixed to equal its transformed item-total biserial correlation, corrected for guessing (3-PLM $\hat{a}$; b - and c - parameters are estimated),

Condition 2: restricted 3-PLM with a fixed to equal 1.0 and the lower asymptote (c) fixed to equal 0.2 for all items (3-PLM $\hat{a}c$; only b parameters are estimated),

Condition 3: restricted 3-PLM with each item's a fixed to equal its transformed item-total biserial correlation, corrected for guessing, and c fixed to equal 0.2 for all items (3-PLM $\hat{a}c$; only b parameters are estimated), and

Condition 4: analytically derived parameter estimates with a fixed to equal its transformed item-total biserial correlation, corrected for guessing; the difficulty (b) parameter fixed to equal its transformed sample-based p -value; and c fixed to equal 0.2 for all items (NoCal; no calibration is implemented and thus no parameters are estimated).

The maximum marginal likelihood (i.e., modal Bayes) estimation procedure for calculating item parameters was implemented using flexMIRT™ version 1.86 (Cai, 2012). Each modeled condition employed a set of prior distributions for parameters estimated using the slope-intercept

parameterization of the 3-PLM, where the slope parameter is equivalent to the a -parameter, and the intercept parameter is a function of both the a - and b -parameters and can be defined as

$$\text{intercept} = -ab.$$

For parameters estimated (i.e., not fixed) in each model, item-specific priors implemented included a log-normal distribution with a mean equal to the transformed sample biserial item-test correlation, corrected for guessing (see Appendix B) and standard deviation of 1.0 for the slope parameter; and a normal distribution with a mean equal to the transformed sample p -value (see Appendix A) and standard deviation of 3.0 for the intercept parameter. More specifically, Condition 0 implements the slope prior; Conditions 0–3 implement the intercept prior.

A general normally distributed prior (i.e., same prior used for all items) was implemented for the c -parameter such that the distribution had a mean equal to approximately -1.39 (i.e., the logit of the c -parameter where the c -parameter mode is equivalent to 0.20) and standard deviation equal to 0.5. Conditions 0–2 and 5 implement this c -parameter prior. Additionally, Condition 5 implements general priors for the item slopes and intercepts, where the item slopes are distributed log-normal with a mean equal to zero and standard deviation equal to one and the item intercepts are distributed normally with a mean equal to zero and standard deviation equal to three.

Assessments and Examinee Sample

Some DLPT5 language assessments (e.g., Spanish Listening) are afforded the luxury of the accumulation of 1,000 or more examinees such that adequate sample sizes are available for calibrating item parameters using the 3-PLM. Examinee data from the Spanish Listening assessment was used to conduct this study. The complete examinee pool was used to acquire “trustworthy” item parameters (i.e., the control condition to which all parameter estimates are compared) and conduct model-fit analyses for each model employed. Parameters are also estimated using random samples drawn from this pool of examinees at sample sizes of 25, 50, 75, 100, 150, 200, and 1,000. Sampling was replicated 100 times for each sample size. The evaluation of seven sample sizes and six models (only five requiring calibration) results in 42 study conditions. Each of the 42 study conditions is replicated 100 times, resulting in a total of 3,505 calibration analyses (including the control condition and conditions used for model fit analyses) and 701 analytically derived solutions. For replications with small sample sizes, a calibration solution did not always converge. In these cases, the replication was dropped and

the total number of replications that converged was recorded and used moving forward in the study. A program was developed for efficiently running each calibration and the subsequent calculations required for the evaluation criteria.

Ability Estimation

For each study condition and replication, a set of item parameter estimates were obtained. These estimates were used to calculate *expected a posteriori* (EAP) ability estimates for summed scores (i.e., number correct scores) using the following formula (Thissen & Orlando, 2001):

$$\hat{\theta}_j = E(\theta | NC_j) = \frac{\int_{-\infty}^{+\infty} L(NC_j | \theta) \phi(\theta) \theta d\theta}{\int_{-\infty}^{+\infty} L(NC_j | \theta) \phi(\theta) d\theta} \cong \frac{\sum_{q=1}^Q L(NC_j | X_q) A(X_q) X_q}{\sum_{q=1}^Q L(NC_j | X_q) A(X_q)},$$

where $\hat{\theta}_j$ is the estimated ability parameter for examinee j , q indexes the quadrature points ($q = 1, 2, 3, \dots, Q$), X_q represents the quadrature points of a Gaussian distribution, L represents the likelihood function, and A represents weights for the prior Gaussian distribution [i.e., $N \sim (0,1)$]. These summed score EAP estimates were used for classifying examinees into ILR proficiency levels based on the transformed number correct cut scores (i.e., number correct cut score transformed to the theta scale) obtained from the standard setting workshop.

Evaluation Criteria

Estimates of root mean squared error (RMSE), bias, and standard error for item parameters were calculated using the control condition parameter estimates as the point of comparison for the conditions relying upon smaller sample sizes to calibrate items. In addition to single parameter stability and accuracy evaluations, item characteristic curve model fit was also evaluated using RMSE and model fit analyses. Finally, classification decision comparisons were conducted.

Parameter Recovery

The accuracy of parameter recovery for each of the 42 study conditions (model $\times$ sample size) was calculated through comparison of the sample-based estimate with the control condition parameter, re-parameterized as a -, b -, and c -parameters. For any given item (i), let τ_{ip} be the true value of the parameter [p ; item parameter (a , b , or c)]; t_{kip} is the empirical estimate of τ_{ip} for the k^{th} replication. The number of replications is R . The RMSE of a parameter estimate (a , b , c) for item i and parameter p is defined as

$$RMSE_{ip} = \sqrt{\sum_{k=1}^R (t_{kip} - \tau_{ip})^2 / R}.$$

The average RMSE across items was reported for each item parameter (a , b , and c) for each study condition. RMSE is the square root of the Mean Square Error (MSE), which has been shown to decompose to the sum of estimation bias and sampling variance (i.e., $MSE = \text{Squared Bias} + \text{Variance}$). Therefore, a large MSE may result from bias in estimation, sampling variance, or both. The square root of the MSE components—squared bias and variance—provide estimates of bias and standard error (SE), respectively and are defined as

$$Bias_{ip} = \bar{t}_{ip} - \tau_{ip}$$

and

$$SE_{ip} = \sqrt{\sum_{k=1}^R (t_{kip} - \bar{t}_{ip})^2 / R}.$$

The RMSE of the estimated Item Characteristic Curve (ICC) for each item are also calculated. Evaluating the RMSE of the ICC rather than each individual item parameter allows for a better understanding of how well the ICC fits the data. This calculation uses 61 different equally spaced (interval = 0.1) θ values from $j = -3$ to $j = 3$. Let $P(\theta_j)$ be the value of the control condition's ICC for each item when $\theta_j = j$ and $\widehat{P}(\theta_j)$ be the sample-based estimate, then the $RMSE_{icc}$ for each item within a study condition is defined as

$$RMSE_{icc} = \sqrt{\sum_{j=1}^{61} (\widehat{P}(\theta_j) - P(\theta_j))^2 / 61}.$$

Finally, for a set of equally spaced known theta values (θ_j), the weighted average squared-distances ($\bar{d}$) between the test characteristic curves (TCCs) were computed for the control and study condition item calibrations, where the weights were the normal probability densities (standardized to sum to unity):

$$\bar{d} = \sum_{j=1}^s \frac{\varphi_j [\widehat{P}(\theta_j) - P(\theta_j)]^2}{\sum_{j=1}^s \varphi_j},$$

where $\widehat{P}(\theta_j)$ is a point on the study condition's TCC, $P(\theta_j)$ is a point on the control condition's TCC, and φ_j is the normal density at θ_j . For each study condition, the item parameters used to compute TCCs were the average of the 100 replications.

Model Fit

Model fit can be assessed in a variety of ways, yet no existing methods are considered best (e.g., Yen & Fitzpatrick, 2006). Several sources (e.g., Sinharay, 2003) have recognized that of the fit statistics widely used in practice, the methods proposed by Orlando and Thissen (2000, 2003) are the most robust to the common challenges (e.g., inflated Type I error rates) posed by statistical tests of item fit. Orlando and Thissen (2000, 2003) adapt the χ^2 (as Q_1 ; Yen, 1981) and G^2 (McKinley & Mills, 1985) statistics such that K groups are formed based on an observed variable (i.e., raw score) instead of forming groups based on the unobserved variable (i.e., estimated ability, theta) which serves to resolve the violation of one of the statistic's assumption in calculating the expected proportions. The modified statistics are defined as

$$\chi_i^2 = \sum_{k=1}^K N_k \frac{(p_{ik} - E_{ik})^2}{E_{ik}(1 - E_{ik})}$$

and

$$G_i^2 = 2 \sum_{k=1}^K N_k \left[p_{ik} \log \left(\frac{p_{ik}}{E_{ik}} \right) + (1 - p_{ik}) \log \left(\frac{1 - p_{ik}}{1 - E_{ik}} \right) \right],$$

where N_k is the number of examinees in group k , p_{ik} is the proportion of examinees responding correctly to item i in group k , and E_{ik} is the proportion of examinees expected to respond correctly to item i in group k based on the model. One group for each possible total score was formed for calculating these statistics using each of the models and the complete data set. For each statistic, the values were summed across items to assess model-data fit at a global level as defined here:

$$\chi_{T(m)}^2 = \sum_{i=1}^n \chi_{i(m)}^2$$

or

$$G_{T(m)}^2 = \sum_{i=1}^n G_{i(m)}^2,$$

where n is equal to the total number of items and m is the acronym (e.g., 3-PLM) for the model fit. Each of the summed statistics are considered to be distributed asymptotically as a chi-square and were evaluated as such with $n(k-p-2)$ degrees of freedom where p is equal to the number of parameters estimated in the model. To statistically compare models (e.g., Does a 3-PLM

significantly fit the data better than the restricted models?), Waller (1981) showed that the difference between the global statistics can be assessed with n degrees of freedom. A non-statistically (N.S.) significant difference would indicate a failure to reject the null hypothesis that the model with more parameters fits the data better or more specifically (i.e., using $\chi^2_{T(m)}$ as the example). For each of the model comparisons proposed the null hypotheses can be stated as

1. $\chi^2_{T(3-PLM,general)} - \chi^2_{T(3-PLM,item-specific)}$; H_0 : all item parameters can be estimated using general priors;
2. $\chi^2_{T(3-PLM\hat{a})} - \chi^2_{T(3-PLM,general)}$; H_0 : $a_i = \hat{a}_i$, all $i = 1, \dots, n$ where $\hat{a}_i$ is equal to the item-specific fixed value;
3. $\chi^2_{T(3-PLM\hat{a}c)} - \chi^2_{T(3-PLM\hat{a})}$; H_0 : $c_i = 0.2$, all $i = 1, \dots, n$;
4. $\chi^2_{T(3-PLMac)} - \chi^2_{T(3-PLM\hat{a})}$; H_0 : $a_i = 1.0$ and $c_i = 0.2$, all $i = 1, \dots, n$;
5. $\chi^2_{T(3-PLMac)} - \chi^2_{T(3-PLM\hat{a}c)}$; H_0 : $a_i = 1.0$, all $i = 1, \dots, n$;
6. $\chi^2_{T(NoCal)} - \chi^2_{T(3-PLM\hat{a}c)}$; H_0 : $b_i = \hat{b}_i$, all $i = 1, \dots, n$ where $\hat{b}_i$ is equal to the item-specific fixed value.

This same set of model comparisons and null hypotheses can be used for the $G^2_{T(m)}$ as well.

These stepwise model comparisons will help determine which model best fits the DLPT5 language assessment data.

Examinee ILR Level Classification

Quantification of bias and variance in parameter recovery and model fit may be of less interest to stakeholders than their impact on examinee classification (i.e., ILR level). Therefore, the potential for misclassification of examinees by ILR level for each of the 42 study conditions was also examined. Classification comparisons were conducted in two ways: 1) decision accuracy based on the modeled number correct scale (DA_{NC}), and 2) decision consistency based on the empirically derived theta scale (DC_θ). DA_{NC} were calculated using the following steps:

1. Calculate the probability of each number correct score (NC_n) given known theta values $[P(NC_n|\theta_q)]$ along equally spaced quadrature points (q).
2. Integrate across all probabilities for each NC to obtain the marginal distribution for each of the NCs.
3. Use the cut scores obtained from standard setting to calculate the probability of being classified in or above each of the ILRs.

4. Complete steps 1–3 for the control and sample conditions.
5. Compare the values obtained in the control condition with the sample conditions.

DC_{θ} were obtained by comparing the classification consistency between examinees in the control and sample conditions as outlined in Table 1. Examinees were classified based on their summed score EAP estimates as described earlier. The sum of the diagonal elements in Table 1 represents DC_{θ} . To control for chance agreement, Cohen's (1960) kappa (κ) coefficient was also calculated and is defined as

$$\kappa = \frac{(Observed\ Agreement) - (Chance\ Agreement)}{1 - (Chance\ Agreement)} = \frac{\sum_l C_{ll} - \sum_l C_{l.}C_{.l}}{1 - \sum_l C_{l.}C_{.l}},$$

where C_{ll} is equal to the proportion of examinees whose observed performance level (l ; $l=0, 0+, 1, 1+, 2, 2+, \text{ or } 3$) is l both within the control and study conditions, $C_{l.}$ is equal to the proportion of examinees whose observed performance level is l for the control condition, and $C_{.l}$ is equal to the proportion of examinees whose observed performance level is l for the study condition. For each study condition, the mean for the κ statistic across replications was reported.

Table 1. Examinee Classification: Control Condition vs. Study Condition X

		Control Condition							Total
		0	0+	1	1+	2	2+	3	
Study Condition X	0	41%							xxx
	0+	41%	0%						xxx
	1	19%	66%	0%					xxx
	1+		34%	48%	0%				xxx
	2			52%	100%	100%			xxx
	2+						100%	25%	xxx
	3							75%	xxx
	Total	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxxx

Results

This section describes the results obtained for each of the evaluation criteria estimated to include calibration failure rates, estimates of root mean squared error (RMSE), bias, and standard error for item parameters, item characteristic curve model fit evaluated using RMSE, model fit analyses, and classification decision accuracy and consistency outcomes for each of the models and sample sizes employed. Each model (i.e., study condition) was assessed based on each of the criterion or set of criteria in turn.

Failed Calibrations

Even while item parameter prior distributions were used for estimating item parameters, maximum marginal likelihood was used as the estimation technique; thus, in the event that the sample drawn from the complete set of response patterns contained examinees for which none or all answered incorrectly or correctly to an item on the assessment, the calibration of item parameter estimates would fail. In such cases, the replication was dropped from the study and evaluation criteria were reported in the absence of these replications. Table 2 reports the number of dropped replications. Reported summary statistics must be interpreted with caution for sample sizes of 25–100 (especially for sample sizes of 25 and 50) as the dropped replications may serve to skew the distribution of examinees.

Table 2. Number of Failed Calibrations out of 100 Replications

Sample Size Condition	# Failed
1000	0
200	0
150	0
100	6
75	11
50	43
25	88

Parameter Recovery

Parameter recovery is described for each study condition and includes RMSE, bias, and standard error for the discrimination, difficulty, and lower asymptote item parameter estimates in this section. Item parameter estimates (treated as true item parameters) for the control condition

are found in Table C.1 of Appendix C. Average (across all items) RMSE, bias, and standard error estimates for item parameter estimates can be found in Table D.1 of Appendix D.

Item-level RMSE, bias, and standard error estimates for item parameter estimates can be found in Tables E.1–E.42 of Appendix E. Average and item-level ICC RMSE can also be found in Tables D.1 and Tables E.1–E.42. Further, weighted distances between control and study TCCs can be found in Table D.1 of Appendix D. Lastly, graphical displays of average TCCs can be found in Figures F.1–F.7 of Appendix F.

Condition 0: Item-Specific Priors, 3-PLM

Average RMSE for the discrimination parameter ranged from 0.202 ($n=1,000$) to 1.191 ($n=25$) and progressively increased as sample size decreased. A sample size of 1,000 was the only condition for which the average RMSE did not exceed 0.40, a reasonable threshold used to judge parameter estimate adequacy. Average bias ranged from 0.022 ($n=150$ and considered negligible) to 0.409 ($n=25$ and considered significant). Average standard error ranged from 0.193 ($n=1,000$) to 1.054 ($n=25$) with standard error increasing as sample size decreased.

Average RMSE for the difficulty parameter ranged from 0.194 ($n=1,000$) to 1.173 ($n=25$) and progressively increased as sample size decreased. A sample size of 1,000 was the only condition for which the average RMSE for the difficulty parameter estimate did not exceed 0.40, a reasonable threshold used to judge parameter estimate adequacy. Average bias was considered negligible for all sample size conditions with the exception that a bias of 0.218 ($n=25$) was considered moderate. Average standard error ranged from 0.172 ($n=1,000$) to 1.063 ($n=25$) with standard error increasing as sample size decreased.

Average RMSE for the lower asymptote parameter ranges from 0.041 ($n=1,000$) to 0.057 ($n=50$). A sample size of 1,000 was the only condition for which the average RMSE for the lower asymptote did not exceed 0.05, a reasonable threshold used to judge parameter estimate adequacy. Average bias ranged from 0.008 ($n=1,000$) to 0.013 ($n=75$), all considered negligible. Average standard error ranged from 0.010 ($n=25$) to 0.024 ($n=1,000$), an unexpected finding as standard error decreased as sample size decreased.

Average RMSE for the conditional probability for responding correctly to an item across the sixty-one quadrature points (i.e., ability levels) ranged from 0.170 ($n=1,000$) to 0.602 ($n=75$) with the only reported value below 0.30 (a reasonable threshold used to judge parameter estimate adequacy) being for a sample size of 1,000. Aggregating item conditional probabilities across

items resulted in expected conditional number correct scores and a test characteristic curve (TCC). The weighted (based on normal curve probits) distance between the control condition TCC and the study condition TCCs ranged from 0.014 (n=50) to 0.511 (n=25). Distances scaled (i.e., divided by the number of items on the test) based on total possible points were not expected to exceed 0.02 as experience with equating studies has shown that larger weighted distances that exist between TCCs results in significant equating error. Therefore, a criterion of 0.02 multiplied by 50 (the number of items on the test) was used as the threshold in this study. Thus, all distances between TCCs were considered acceptable as none exceeded 1.00 on the number correct scale.

Condition 1: Item-Specific Priors, 3-PLM $\hat{\mu}$

Average RMSE for the discrimination parameter ranged from 0.423 (n=1,000) to 0.710 (n=25) and progressively increased as sample size decreased. No sample size condition resulted where RMSE was less than 0.40. Average bias ranged from -0.377 (n=1,000) to -0.180 (n=25), an unexpected finding as the absolute magnitude of bias increased with sample size. Average standard error ranged from 0.114 (n=1,000) to 0.563 (n=25) with standard error increasing as sample size decreased, as expected.

Average RMSE for the difficulty parameter ranged from 0.438 (n=1,000) to 2.771 (n=25) and progressively increased as sample size decreased. No sample size condition resulted where RMSE was less than 0.40. The smallest average bias magnitude observed was -0.019 (n=100 and considered negligible) and the largest in magnitude was 0.823 (n=25 and considered very large). Average standard error ranged from 0.214 (n=1,000) to 2.586 (n=50) with standard error generally increasing as sample size decreased.

Average RMSE for the lower asymptote parameter ranges from 0.046 (n=1,000) to 0.057 (n=25). A sample size of 1,000 was the only condition for which the average RMSE for the lower asymptote did not exceed 0.05. The smallest average bias was observed for a sample size of 200 at a magnitude of 0.001 and the largest for a sample size of 1,000 at a magnitude of -0.016, all of which were considered negligible. Average standard error ranged from 0.010 (n=25) to 0.024 (n=1,000), an unexpected finding as standard error decreased as sample size decreased.

Average RMSE for the conditional probability for responding correctly to an item across the sixty-one quadrature points (i.e., ability levels) ranged from 0.354 (n=25) to 0.755 (n=75)

with no sample size resulting in a value below 0.30. The weighted distance between the control condition and the study condition TCCs ranged from 1.242 (n=75) to 27.920 (n=25). Even while the average difference between item characteristic curves was reported as the lowest for a sample size of 25, this sample size resulted in the largest difference between TCCs, likely due to the significantly greater average bias observed for the difficulty parameter accumulating across items. Overall, all distances between TCCs were considered unacceptable as all exceeded 1.00.

Condition 2: Item-Specific Priors, 3-PLMac

Average RMSE for the discrimination parameter was 0.635 for all sample sizes as this parameter was fixed to equal 1.0. Likewise, since the average discrimination parameter in the control sample was 1.491, the average bias for all sample size conditions was -0.491 with no standard error observed across the 100 replications, as expected.

Average RMSE for the difficulty parameter ranged from 0.594 (n=1,000) to 0.914 (n=25) and progressively increased as sample size decreased. No sample size condition resulted where RMSE was less than 0.40. The smallest average bias magnitude observed was -0.057 (n=50 and considered negligible), and the largest in magnitude was -0.241 (n=1,000 and considered moderate). Average standard error ranged from 0.105 (n=1,000) to 0.693 (n=25) with standard error increasing as sample size decreased.

As with the discrimination parameter, the lower asymptote was fixed to a constant value (0.20); thus, the average RMSE was 0.058 for all sample sizes. Likewise, since the average lower asymptote parameter in the control sample was 0.186, the average bias for all sample size conditions was 0.014 with no standard error observed across the 100 replications, as expected.

Average RMSE for the conditional probability for responding correctly to an item across the 61 quadrature points ranged from 0.286 (n=25) to 0.660 (n=75) with only a sample size of 25 resulting in a value below 0.30, an unexpected finding. However, the positive bias observed in the difficulty parameter for a sample size of 25 accumulated across items such that the weighted distance between the control condition and the study condition TCCs was largest ($\bar{d} = 4.646$) for this sample size. Predictably, the lowest distance of 1.990 was observed for a sample size of 1,000. Still, all distances between TCCs were considered unacceptable as all exceeded 1.00.

Condition 3: Item-Specific Priors, 3-PLMâc

Average RMSE for the discrimination parameter ranged from 0.423 (n=1,000) to 0.710 (n=25) and progressively increased as sample size decreased. No sample size condition resulted where RMSE was less than 0.40. Average bias ranged from -0.377 (n=1,000) to -0.180 (n=25), an unexpected finding as the absolute magnitude of bias increases with sample size. Average standard error ranged from 0.114 (n=1,000) to 0.563 (n=25) with standard error increasing as sample size decreased, as expected.

Average RMSE for the difficulty parameter ranged from 0.453 (n=1,000) to 2.788 (n=25) and progressively increased as sample size decreased. No sample size condition resulted where RMSE was less than 0.40. The smallest average bias magnitude observed was 0.008 (n=100 and considered negligible), and the largest in magnitude was 0.832 (n=25 and considered very large). Average standard error ranged from 0.203 (n=1,000) to 2.594 (n=50) with standard error generally increasing as sample size decreased.

Average RMSE for the lower asymptote parameter was 0.058 for all sample sizes, as this parameter was fixed to equal 0.20. Likewise, since the average lower asymptote parameter in the control sample was 0.186, the average bias for all sample size conditions was 0.014 with no standard error observed across the 100 replications, as expected.

Average RMSE for the conditional probability for responding correctly to an item across the 61 quadrature points (i.e., ability levels) ranged from 0.352 (n=25) to 0.759 (n=75) with no sample size resulting in a value below 0.30. The weighted distance between the control condition and the study condition TCCs ranged from 1.271 (n=75) to 26.850 (n=25). Even while the average difference between item characteristic curves was reported as the lowest for a sample size of 25, this sample size resulted in the largest difference between TCCs, likely due to the significantly greater average bias observed for the difficulty parameter accumulating across items. Overall, all distances between TCCs were considered unacceptable as all exceeded 1.00.

Condition 4: Analytically Derived Parameter Estimates, NoCal

Average RMSE for the discrimination parameter ranged from 0.423 (n=1,000) to 0.710 (n=25) and progressively increased as sample size decreased. No sample size condition resulted where RMSE was less than 0.40. Average bias ranged from -0.377 (n=1,000) to -0.180 (n=25), an unexpected finding as the absolute magnitude of bias increased with sample size. Average

standard error ranged from 0.114 ($n=1,000$) to 0.563 ($n=25$) with standard error increasing as sample size decreased, as expected.

Average RMSE for the difficulty parameter ranged from 0.292 ($n=1,000$) to 1.117 ($n=25$) and progressively increased as sample size decreased. Only a sample size of 1,000 resulted where RMSE was less than 0.40. The smallest average bias magnitude observed was 0.080 ($n=75$ and considered negligible), and the largest in magnitude was 0.265 ($n=25$ and considered moderate). Average standard error ranged from 0.139 ($n=1,000$) to 0.944 ($n=25$) with standard error increasing as sample size decreased.

Average RMSE for the lower asymptote parameter was 0.058 for all sample sizes, as this parameter was fixed to equal 0.20. Likewise, since the average lower asymptote parameter in the control sample was 0.186, the average bias for all sample size conditions was 0.014 with no standard error observed across the 100 replications, as expected.

Average RMSE for the conditional probability for responding correctly to an item across the 61 quadrature points ranged from 0.310 ($n=25$) to 0.615 ($n=75$) with no sample size condition resulting in a value below 0.30. The positive and moderate bias observed in the difficulty parameter for a sample sizes of 25, 150, 200, and 1,000 accumulated across items such that the weighted distance between the control condition and the study condition TCCs were larger than 1.00 for each of these conditions. For sample sizes of 50–100, the weighted distance between TCCs all fell below the acceptable threshold of 1.00.

Condition 5: General Priors, 3-PLM

Average RMSE for the discrimination parameter ranged from 0.189 ($n=1,000$) to 0.790 ($n=25$) and progressively increased as sample size decreased. A sample size of 1,000 was the only condition for which the average RMSE did not exceed 0.40. Average bias ranged from -0.048 ($n=1,000$ and considered negligible) to -0.626 ($n=25$ and considered significant). Average standard error ranged from 0.173 ($n=1,000$) to 0.431 ($n=75$).

Average RMSE for the difficulty parameter ranged from 0.201 ($n=1,000$) to 1.570 ($n=25$) and progressively increased as sample size decreased. A sample size of 1,000 was the only condition for which the average RMSE for the difficulty parameter estimate did not exceed 0.40. The least average bias observed was 0.000 ($n=1,000$ and considered non-existent), and the largest in magnitude was -0.219 ($n=75$ and considered moderate). Average standard error ranged from 0.178 ($n=1,000$) to 1.315 ($n=25$) with standard error increasing as sample size decreased.

Average RMSE for the lower asymptote parameter ranged from 0.040 ($n=1,000$) to 0.056 ($n=25$). A sample size of 1,000 was the only condition for which the average RMSE for the lower asymptote did not exceed 0.05. Average bias ranged from 0.005 ($n=1,000$) to 0.010 ($n=75$), all considered negligible. Average standard error ranged from 0.008 ($n=25$) to 0.022 ($n=1,000$), an unexpected finding as standard error decreased as sample size decreased.

Average RMSE for the conditional probability for responding correctly to an item across the 61 quadrature points ranged from 0.174 ($n=1,000$) to 0.680 ($n=75$) with the only reported value below 0.30 being for a sample size of 1,000. The weighted distance between the control condition and the study condition TCCs ranged from 0.027 ($n=1,000$) to 5.882 ($n=25$) with only sample size conditions of 200 and 1,000 reporting distances below the threshold of 1.00.

Summary

Of all the conditions modeled, Condition 0, where item-specific priors are employed for the estimation of the intercept and slope parameters, seems to result in the best fit (measured using a weighted distance formula) between study and control TCCs, especially in the presence of small sample sizes (200 or smaller). At the level of the ICC, it appears again that Condition 0 consistently results in the lowest average RMSE across sample size conditions. At the level of individual parameters, stating a condition that best recovers the parameter in question is not as clear, especially as sample size decreases. However, it was observed that the average evaluation criterion most highly correlated with the weighted distance between control and study TCCs was absolute average bias of the difficulty parameter with a correlation of 0.88, suggesting that the extent to which bias in estimating difficulty can be controlled will result in better estimates of ability for examinees.

Condition 1, where slopes were fixed based on empirical estimation and intercepts and lower asymptotes were estimated, did not work sufficiently to suggest its use as all reproduced distances between TCCs were greater than 1.00. While fixing slopes equal to one and fixing the lower asymptote to 0.20 across all items and estimating difficulty (Condition 2) worked better than Condition 1 to reduce distances between TCCs, it still was not sufficient in comparison to the threshold of 1.00 and in comparison to Condition 0. Likewise, fixing slopes based on empirical estimation and fixing lower asymptotes to 0.20 and estimating difficulty (Condition 3) worked no better than Condition 1 and performed worse than Condition 2. Under small sample conditions (25–200), Condition 4 (where all parameters are empirically estimated) worked better

than Conditions 1–3, but did not perform better than Condition 0, which makes sense since it is the Condition 4 parameter estimates that serve as the prior means during estimation using Condition 0. Finally, Condition 5, where general priors across all items are used, served to perform comparably to Condition 0 for a sample size of 1,000. However, under smaller sample sizes (25–100), Condition 5 did not outperform either Condition 0 or Condition 4.

Model Fit

Table 3 displays both χ^2 and G^2 values for each model along with the difference between selected models. As these statistics are sensitive to large sample sizes, all models show misfit. Therefore, it is more informative to look at the change in these values between models. Evaluated with fifty degrees of freedom (the number of items on the test), it appears that in the full set of data, both models that estimate all parameters fit the data best and are seemingly indistinguishable in terms of fit. All other models are significantly different from one another; that is, fixing item slopes to estimated values produces significantly worse fit than estimating the slope parameter using either general or item-specific priors. Worse still is the fit associated with fixing both slope and lower asymptote parameter estimates (Conditions 2 and 3). Finally, fixing all parameter values results in the worst fit among all the models. Overall, both χ^2 and G^2 produced consistent results.

Table 3. Model Fit Diagnostic Statistics, χ^2 and G^2

Model	χ^2 *	G^2 *	d.f.
C0	2927.04	2600.76	1850
C1	5641.68	5280.18	1900
C2	7539.21	7022.83	1950
C3	6449.96	6102.68	1950
C4	8592.87	9414.33	2000
C5	2936.89	2621.00	1850
C5–C0	9.85	20.24	50.00
C1–C5	2704.79	2659.18	50.00
C3–C1	808.28	822.50	50.00
C2–C1	1897.53	1742.65	50.00
C2–C3	1089.25	920.15	50.00
C4–C3	2142.91	3311.65	50.00

* A significant change in χ^2 or G^2 is equal to 67.51.

Classification

Classification accuracy and consistency into ILR levels along with score reliability is described for each study condition. Classification accuracy and reliability estimates can be found in Tables G.1–G.7 of Appendix G. Classification consistency and kappa estimates can be found in Tables H.1–H.42 of Appendix H.

Based on the modeled probabilities in the complete set of data ($n=4,187$), 0.2% of examinees were expected to score at a level zero, 0.3% at a level zero plus, 4.0% at a level one, 9.6% at a level one plus, 22.7% at a level two, 22.7% at a level two plus, and 40.5% at level three. Scores for examinees in the complete set of data reveal 0.0% classified as level zero, 0.2% at level zero plus, 4.5% at level one, 6.8% at level one plus, 22.2% at level two, 31.9% at level two plus, and 34.4% at level three. To the extent that error exists in the model, reliability will be weakened. For the modeled probabilities, the reliability is estimated to be 0.85. Modeled classification probabilities are used for comparison to assess classification accuracy, while observed classification percentages are used for comparison to assess classification consistency. The sum of modeled errors for the complete set of data is 19.3%, which is spread across seven classification levels. Thus, the average modeled error within a level was 2.8% and had a standard deviation of 3.5. Therefore, the threshold used for assessing acceptable decision consistency across all levels was 77% (100% – sum of modeled error plus one standard deviation).

Condition 0: Item-Specific Priors, 3-PLM

The difference between modeled classification expectations between the study and control conditions ranged from a low of 1.7% ($n=1,000$) to 19.4% ($n=25$) with the difference increasing as sample size decreased. Differences in reliability ranged from 0.003 ($n=1,000$) to 0.013 ($n=25$) with no clear pattern associated with sample size. Both decision consistency and kappa values decreased with decreasing sample size with highs of 94.06% for decision consistency and 91.78 for kappa with a sample size of 1,000 and lows of 70.00% for decision consistency and 57.81 for kappa with a sample size of 25. All sample size conditions met the threshold of 77% decision consistency except for the sample size of 25.

Condition 1: Item-Specific Priors, 3-PLM $\hat{\mu}$

The difference between modeled classification expectations between the study and control conditions ranged from a low of 15.4% ($n=1,000$) to 23.5% ($n=50$) with the difference generally increasing as sample size decreased. Differences in reliability ranged from -0.033

($n=25$) to -0.076 ($n=150$) with differences generally decreasing in magnitude as sample size decreased—an unexpected finding. Both decision consistency and kappa values decreased with decreasing sample size with highs of 86.34% for decision consistency and 81.24 for kappa with a sample size of 1,000 and lows of 65.00% for decision consistency and 51.42 for kappa with a sample size of 25. All sample size conditions met the threshold of 77% decision consistency except for the sample sizes of 25 and 50.

Condition 2: Item-Specific Priors, 3-PLMac

The difference between modeled classification expectations between the study and control conditions ranged from a low of 17.1% ($n=1,000$) to 34.5% ($n=25$) with the difference increasing as sample size decreased. Differences in reliability ranged from -0.027 ($n=75$) to -0.084 ($n=1,000$) with no clear association with differences and sample size. Both decision consistency and kappa values decreased with decreasing sample size with highs of 85.39% for decision consistency and 79.81 for kappa with a sample size of 1,000 and lows of 66.00% for decision consistency and 52.46 for kappa with a sample size of 25. All sample size conditions met the threshold of 77% decision consistency except for the sample sizes of 25 and 50.

Condition 3: Item-Specific Priors, 3-PLMac

The difference between modeled classification expectations between the study and control conditions ranged from a low of 17.1% ($n=1,000$) to 23.6% ($n=50$) with the difference generally increasing as sample size decreased. Differences in reliability ranged from -0.037 ($n=25$) to -0.084 ($n=1,000$) with differences decreasing in magnitude as sample size decreased, an unexpected finding. Both decision consistency and kappa values decreased with decreasing sample size with highs of 84.88% for decision consistency and 79.27 for kappa with a sample size of 1,000 and lows of 66.00% for decision consistency and 52.80 for kappa with a sample size of 25. All sample size conditions met the threshold of 77% decision consistency except for the sample sizes of 25 and 50.

Condition 4: Analytically Derived Parameter Estimates, NoCal

The difference between modeled classification expectations between the study and control conditions ranged from 21.5% ($n=1,000$) to 27.3% ($n=25$) with the difference generally increasing as sample size decreased. Differences in reliability ranged from -0.017 ($n=25$) to -0.046 ($n=150$) with no clear association between differences and sample size. Both decision consistency and kappa values decreased with decreasing sample size with highs of 75.80% for

decision consistency and 65.92 for kappa with a sample size of 1,000 and lows of 63.67% for decision consistency and 49.60 for kappa with a sample size of 25. No sample size condition met the threshold of 77% decision consistency.

Condition 5: General Priors, 3-PLM

The difference between modeled classification expectations between the study and control conditions ranged from 3.1% (n=1,000) to 48.8% (n=25) with the difference increasing as sample size decreased. Differences in reliability ranged from -0.003 (n=1,000) to -0.087 (n=25) with the magnitude of the difference increasing with decreasing sample size. Both decision consistency and kappa values decreased with decreasing sample size with highs of 92.40% for decision consistency and 89.47 for kappa with a sample size of 1,000 and lows of 60.00% for decision consistency and 45.09 for kappa with a sample size of 25. Sample sizes of 100–1,000 met the threshold of 77% decision consistency.

Classification Differences between Two Groups

Focusing on a single cut score, scoring at level two or higher, classification decisions can be analyzed differently. Due to the collapsing of scoring categories, it is expected that fewer examinees would be differently classified into each of the two scoring groups (below level two and level two or higher). Table 4 displays classification differences observed using the differing models under the various sample size conditions. The suggested acceptable classification difference is one percent. However, in interpreting results, it is important to remember that not all replications were included in sample size conditions 25–100 due to failed calibration attempts. Therefore, the included replications may produce different sample summary statistics than expected, which could limit the comparability between sample size conditions. Therefore, results with sample sizes of 100 or fewer should be interpreted with caution.

Table 4. Percent Difference of Classification in Level 2 or Higher between Condition and Control, based on Observed Scores

Condition	1000	200	150	100	75	50	25
0	-0.19	0.06	0.31	0.67	0.82	2.53	4.00
1	-3.45	-3.00	-2.87	-2.52	-2.07	-0.35	1.67
2	-2.04	-1.72	-1.63	-1.24	-1.12	0.04	1.67
3	-4.00	-3.64	-3.09	-2.70	-2.23	-0.63	1.33
4	1.53	1.59	1.76	1.83	1.95	3.44	4.00
5	-0.39	-0.64	-0.81	-0.76	-0.84	-0.11	1.67

The best sample size to use for choosing a viable calibration method is 1,000 as it is the largest. Therefore, of all calibration methods, Conditions 0 and 5 resulted in the fewest classification differences between control and study conditions. Within Condition 0, sample sizes of 75–1,000 resulted in differences less than one; and within Condition 5, sample sizes of 50–1,000 resulted in differences less than one.

Summary

Of all conditions modeled and under all sample sizes, reliability differences never exceeded 0.1 between control and study conditions. However, while modeled reliability variations were minimal, differences in decision consistency were apparent such that in general larger sample sizes resulted in higher decision consistency, as expected. Further, even among models implemented with a large sample size of 1,000, significant differences between decision consistency existed with Condition 0 resulting in the highest classification consistency of about 94%. In fact, all models met the 77% threshold for classification consistency at a sample size of 1,000 except for Condition 4 where only 76% of examinees were classified consistently between the control and study condition. Condition 5 resulted in about 92% decision consistency and Conditions 1–3 resulted in approximately 85–86% decision consistency. However, performance of models in terms of decision consistency did depend on the sample size employed. For example, for a sample size of 100, while Condition 0 still outperformed the others with a decision consistency rate of 87%, Condition 5 only resulted in a rate of about 75% with most other models (Conditions 1–3) resulting in higher classification consistency rates (80–83%). Condition 4 still underperformed with a consistency rate of 74%. For sample sizes lower than 100 (25–75), models resulting in acceptable (77% or better) decision consistency rates include Conditions 0–3 for a sample size of 75 and Condition 0 for a sample size of 50. No model worked well for a sample size of 25. In fact, for all sample sizes, decision consistency was greatest when Condition 0 was employed for calibration of item parameter estimates, except when only one cut score was employed. Thus, the number of cut scores, placement of cut scores, and distribution of examinees may serve to effect decision consistency enough to suggest the use of different calibration models given what is valued most (e.g., placement into broad or narrow classifications).

Discussion

This section serves to further summarize findings and discuss their impact on the specific operational practice of field testing multiple-choice items. More specifically, in light of all the criteria outcomes presented, recommendations are made in regard to field test sample sizes and item calibration processes. Limitations of this study are also presented along with suggestions for future analyses and studies.

First, it is important to not overlook the number of failed calibrations this study highlighted at sample sizes of 25–100, ranging from a low of 6% ($n=100$) to a high of 88% ($n=25$). For this study, it was fortunate that a larger pool of examinees existed from which to resample in the event that a replication failed. However, in practice, the only sample that will exist is the one selected and, in the event that sample contains examinees that result in estimated p -values of zero or one for one or more items, it will determine whether the item set can be calibrated. It too is true that items may be calibrated by forcing parameter estimates for problematic items or by dropping the items in question; however, it is more likely the case that the sample size is not sufficient in size or heterogeneity to accurately estimate parameters than there is an inherent flaw with the item(s). Therefore, in light of these findings alone, the recommended sample size for field testing items should be no smaller than 100 examinees for a test consisting of similarly functioning items used to assess a population of examinees with similar ability distribution.

Of the models studied (Conditions 0–5), using prior information (based on classical analyses of test data) about both difficulty (p -values) and discrimination (item-total biserial correlations corrected for guessing) of items enhanced the accuracy to which item parameter estimates were recovered (i.e., applying Condition 0). However, these item parameter estimates were still considered unacceptable for all sample sizes less than 1,000. Therefore, it is not recommended that individual item parameter estimates obtained using less than 1,000 examinees be used to assess item quality. For example, it would not be appropriate to use the difficulty index (i.e., b -parameter estimate) to exclude items from a test form due to the mismatch of difficulty and ILR level designation. It too would be inappropriate to use item parameter estimates for constructing computer adaptive test (CAT) item banks. More examinee response data would need to be collected, and item recalibration would need to take place before items could be used in a CAT testing environment.

While item parameter estimates and resulting item characteristic curves were not recovered adequately using small sample sizes, test characteristic curves were recovered with moderate accuracy under small sample conditions. Of all the conditions modeled, Condition 0, where item-specific priors were employed for the estimation of the intercept and slope parameters, resulted in the best fit (measured using a weighted distance formula) between study and control TCCs. Further, in evaluating model fit using the full set of data, the conclusion was drawn that, while Condition 0 did not significantly improve fit over Condition 5, it did fit significantly better than all other conditions (1–4). Employing Condition 0 for calibration resulted in all sample size conditions meeting the 1.00 weighted difference threshold between control and study TCCs. Therefore, using item parameter estimates (even though not adequately recovered) obtained using Condition 0 results in no more than a one point difference on the number correct scoring scale between control and study conditions when item parameter estimates are used in aggregate across all items to model proficiency on a 50-point scale. It is recommended that in the presence of small samples, Condition 0 be employed for calibrating item parameter estimates as setting item-specific priors on both slope and intercept parameters and a general prior on the lower asymptote parameter served to improve TCC recovery.

Regardless of parameter and summary criteria recovery, it needs to be acknowledged that after collection of field test data and creation of test forms, the item parameter estimates calibrated from the field test data will be used to score examinees responding to items on the test forms. Therefore, it could be argued that the most important criterion to consider in making sample size and model recommendations is the classification consistency of examinees into ILR levels as the decisions made based on assigned ILR levels are high stakes in nature and come with potentially high costs and drastic consequences for errors made.

In summary, for all sample sizes, decision consistency was greatest when Condition 0 was employed for calibration, except in relatively few conditions, and was evaluated as acceptable at a decision consistency rate of 77% for sample sizes as low as 50. However, due to the number of failed calibrations at a sample size this low, it is not recommended. It too must be noted that the number of cut scores, placement of cut scores, and distribution of examinees may serve to effect decision consistency enough to suggest the use of different calibration models given what is valued most (e.g., placement into broad or narrow classifications). With this in mind, it is tentatively

recommended that Condition 0 be used to calibrate item parameter estimates for sample sizes of 100 or more examinees for the sole purpose of constructing parallel forms.

Generalizability of the findings in this study may be limited as only one set of assessment data was explored. Therefore, differences in the distribution of item difficulty, examinee ability, and cut scores could impact the recommendations made for operational practice. Future studies should make an attempt to include conditions that vary these distributions along with test length. Applying these recommendations using small samples to other test forms, skills, and language assessments with high examinee volume could serve to cross-validate this study's findings.

References

- Cai, L. (2012). *flexMIRT™ version 1.86: A numerical engine for multilevel item factor analysis and test scoring*. [Computer software]. Seattle, WA: Vector Psychometric Group.
- Cizek, G. J., & Plake, B. S. (2009). *Developing a plan for setting proficiency standards for the DLPT5 examinations* (Tech. Rep. No. Contract M67004-05-D-0009). Defense Language Office (Personnel and Readiness), U. S. Department of Defense.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- McKinley, L. R., & Mills, C. N. (1985). A comparison of several goodness-of-fit statistics. *Applied Psychological Measurement*, 9(1), 49–57.
- Muraki, E., & Bock, R. D. (2003). *PARSCALE 4: IRT item analysis and test scoring for rating-scale data* [Computer software]. Chicago, IL: Scientific Software International.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50–64.
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of the s_x^2 : An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27(4), 289–298.
- Sinharay, S. (2003). *Bayesian item fit analysis for dichotomous item response theory models* (Educational Testing Service RR-03-34). Princeton, NJ.
- Sukin, T. M., & Nicewander, W. A. (2012). *Calibrating item parameters using small samples: A review of the literature and preliminary recommendations* (Pacific Metrics Corporation Research Report No. 12-1). Monterey, CA.
- Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds), *Test Scoring* (pp. 73–140). Mahwah, NH: Lawrence Erlbaum Associates.
- Tsutakawa, R. K., & Johnson, J. C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*, 55(2), 371–390.
- Waller, M. I. (1981). A procedure for comparing logistic latent trait models. *Journal of Educational Measurement*, 18(2), 119–125.
- Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, 5(2), 245–262.
- Yen, W., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111–153). Westport, CT: Praeger Publishers.

Appendix A

Model Definitions

All of the item response theory (IRT) models proposed in this study are based on the 3-PLM (*Condition 0 and Condition 5*), which is defined as

$$P(u_i = 1|a_i, b_i, c_i, \theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}},$$

where $P(u_i = 1|a_i, b_i, c_i, \theta)$ is the probability of an examinee with ability equal to θ responding correctly ($u_i = 1$) to an item (i) with parameters a_i , b_i , and c_i . Parameters a_i describe the discrimination, b_i describe the difficulty, and c_i describe the lower asymptote for i . The remaining conditions modeled are defined below.

Condition 1

3-PLM $\hat{a}$ is defined as

$$P(u_i = 1|\hat{a}_i, b_i, c_i, \theta) = c_i + (1 - c_i) \frac{e^{\hat{a}_i(\theta - b_i)}}{1 + e^{\hat{a}_i(\theta - b_i)}},$$

where $\hat{a}_i$ is the transformed item-total biserial correlation ($\hat{\rho}$) as derived by Lord (1980, p. 31),

$$\hat{a} = \frac{\hat{\rho}}{\sqrt{1 - \hat{\rho}^2}},$$

and were corrected for guessing (see Appendix B).

Condition 2

3-PLMac is defined as

$$P(u_i = 1|b_i, \theta) = 0.2 + 0.8 \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}.$$

Condition 3

3-PLM $\hat{a}c$ is defined as

$$P(u_i = 1|\hat{a}_i, b_i, \theta) = 0.2 + 0.8 \frac{e^{\hat{a}_i(\theta - b_i)}}{1 + e^{\hat{a}_i(\theta - b_i)}}.$$

Condition 4

The NoCal method is derived using

$$P(u_i = 1 | \hat{a}_i, \hat{b}_i, \theta) = 0.2 + 0.8 \frac{e^{\hat{a}_i(\theta - \hat{b}_i)}}{1 + e^{\hat{a}_i(\theta - \hat{b}_i)}},$$

where $\hat{a}_i$ is defined above and $\hat{b}_i$ is defined as

$$\hat{b} = -\phi^{-1}(\hat{p}),$$

where ϕ^{-1} is equal to the inverse of the standard normal cumulative distribution (Lord, 1980, p. 31) and $\hat{p}$ is the sample-based p -value.

Appendix B

Correcting the Biserial Item-Test Correlation for Guessing

The corrected item-test biserial correlation is defined as

$$\rho_{bis}^* = \frac{\rho_{pb}^* \sqrt{p^*(1-p^*)}}{\phi[\Phi^{-1}(p^*)]},$$

where p^* is the guessing-corrected p -value $[(p - c)/(1 - c)]$, $\phi[.]$ is the height of the ordinate of the normal curve at the probit of the corrected p -value, p^* , and ρ_{pb}^* is the corrected point biserial.

The corrected point biserial is defined as

$$\rho_{pb}^*(u_i, y) = \rho_{pb}(u_i, y) \sqrt{\frac{p_i}{(p_i - c_i)}},$$

where ρ_{pb} is the uncorrected point biserial, p is the item's p -value, and c is the guessing probability.

Appendix C

Control Item Parameters

Table C.1. Item Parameter Estimates for the Control Condition, Sample Size = 4187

Item ID	a	b	c
1	1.240	-1.750	0.130
2	1.090	-0.290	0.330
3	1.370	-1.880	0.140
4	1.270	-1.950	0.150
5	0.790	-2.940	0.180
6	1.260	-2.530	0.180
7	1.670	-1.380	0.140
8	2.720	-1.250	0.130
9	3.650	-1.450	0.310
10	1.640	-2.210	0.170
11	1.850	-1.410	0.170
12	1.710	-2.130	0.290
13	1.510	-1.980	0.190
14	1.780	-1.000	0.310
15	0.770	-2.120	0.150
16	1.180	-0.930	0.150
17	1.980	-2.440	0.220
18	1.800	-1.540	0.170
19	1.270	-0.580	0.310
20	0.260	-3.820	0.200
21	1.280	-0.440	0.120
22	1.690	-1.230	0.120
23	1.390	0.260	0.240
24	0.970	-1.870	0.160
25	2.220	-1.670	0.190
26	0.990	-0.880	0.220
27	1.660	-1.080	0.160
28	0.900	0.460	0.140
29	2.560	-1.960	0.140
30	1.540	-1.070	0.090
31	1.930	-0.720	0.060
32	1.970	-0.230	0.270
33	1.680	-0.540	0.070
34	2.610	-1.750	0.170
35	2.460	-1.810	0.150
36	2.130	1.440	0.160
37	0.510	-0.780	0.210
38	1.020	-1.060	0.190
39	0.960	1.600	0.230
40	1.350	-0.280	0.280
41	0.850	-1.950	0.180
42	2.040	-1.240	0.230
43	1.150	0.620	0.340
44	0.600	-2.030	0.200
45	2.800	-1.430	0.290
46	1.700	-0.080	0.080
47	0.800	1.840	0.120
48	0.900	-0.570	0.110
49	0.790	-0.330	0.130
50	0.310	1.440	0.210

Appendix D
Evaluation Criteria Summary Statistics

Table D.1. Evaluation Criteria Summary Statistics, by Condition and Sample Size

Condition	Sample Size	a_AveRMSE	a_AveBias	a_AveSE	b_AveRMSE	b_AveBias	b_AveSE	c_AveRMSE	c_AveBias	c_AveSE	ICC_AveRMSE	d	d/NumItems
0	1000	0.202	0.029	0.193	0.194	0.031	0.172	0.041	0.008	0.024	0.170	0.020	0.000
1	1000	0.423	-0.377	0.114	0.438	-0.253	0.214	0.046	-0.016	0.018	0.390	2.262	0.045
2	1000	0.635	-0.491	0.000	0.594	-0.241	0.105	0.058	0.014	0.000	0.586	1.990	0.040
3	1000	0.423	-0.377	0.114	0.453	-0.174	0.203	0.058	0.014	0.000	0.446	2.996	0.060
4	1000	0.423	-0.377	0.113	0.292	0.202	0.139	0.058	0.014	0.000	0.449	2.054	0.041
5	1000	0.189	-0.048	0.173	0.201	0.000	0.178	0.040	0.005	0.022	0.174	0.027	0.001
0	200	0.445	0.023	0.425	0.469	0.013	0.428	0.053	0.009	0.016	0.381	0.019	0.000
1	200	0.506	-0.376	0.272	0.811	-0.228	0.651	0.053	0.001	0.011	0.525	2.842	0.057
2	200	0.635	-0.491	0.000	0.667	-0.206	0.281	0.058	0.014	0.000	0.625	2.040	0.041
3	200	0.506	-0.376	0.272	0.840	-0.181	0.657	0.058	0.014	0.000	0.544	3.332	0.067
4	200	0.506	-0.376	0.272	0.469	0.165	0.376	0.058	0.014	0.000	0.505	1.451	0.029
5	200	0.431	-0.247	0.324	0.589	-0.103	0.506	0.053	0.006	0.014	0.425	0.733	0.015
0	150	0.509	0.022	0.487	0.607	0.012	0.562	0.054	0.010	0.015	0.443	0.045	0.001
1	150	0.532	-0.374	0.314	1.160	-0.173	1.024	0.054	0.003	0.010	0.590	2.786	0.056
2	150	0.635	-0.491	0.000	0.703	-0.203	0.334	0.058	0.014	0.000	0.644	2.091	0.042
3	150	0.532	-0.374	0.314	1.182	-0.136	1.030	0.058	0.014	0.000	0.603	3.212	0.064
4	150	0.532	-0.374	0.314	0.561	0.145	0.486	0.058	0.014	0.000	0.536	1.211	0.024
5	150	0.493	-0.305	0.352	0.790	-0.129	0.691	0.054	0.006	0.013	0.508	1.172	0.023
0	100	0.627	0.067	0.603	0.760	-0.011	0.712	0.056	0.012	0.014	0.513	0.139	0.003
1	100	0.571	-0.349	0.393	1.658	-0.019	1.540	0.055	0.007	0.013	0.683	1.889	0.038
2	100	0.635	-0.491	0.000	0.754	-0.189	0.412	0.058	0.014	0.000	0.655	2.115	0.042
3	100	0.571	-0.349	0.393	1.676	0.008	1.543	0.058	0.014	0.000	0.691	2.151	0.043
4	100	0.571	-0.349	0.393	0.744	0.109	0.680	0.058	0.014	0.000	0.572	0.829	0.017
5	100	0.579	-0.371	0.407	1.035	-0.210	0.891	0.055	0.009	0.013	0.589	2.114	0.042
0	75	0.727	0.087	0.701	0.953	-0.017	0.909	0.057	0.013	0.014	0.602	0.229	0.005
1	75	0.611	-0.342	0.448	2.306	0.110	2.201	0.056	0.009	0.013	0.755	1.242	0.025
2	75	0.635	-0.491	0.000	0.786	-0.163	0.471	0.058	0.014	0.000	0.660	2.143	0.043
3	75	0.611	-0.342	0.448	2.325	0.132	2.210	0.058	0.014	0.000	0.759	1.271	0.025
4	75	0.611	-0.342	0.448	0.879	0.080	0.825	0.058	0.014	0.000	0.615	0.685	0.014
5	75	0.645	-0.438	0.431	1.323	-0.219	1.180	0.055	0.010	0.013	0.680	2.663	0.053
0	50	0.873	0.151	0.831	1.050	0.076	0.988	0.057	0.011	0.012	0.537	0.014	0.000
1	50	0.644	-0.316	0.486	2.711	0.424	2.586	0.056	0.008	0.009	0.653	5.874	0.118
2	50	0.635	-0.491	0.000	0.779	-0.057	0.517	0.058	0.014	0.000	0.534	2.871	0.057
3	50	0.644	-0.316	0.486	2.725	0.439	2.594	0.058	0.014	0.000	0.653	5.414	0.108
4	50	0.644	-0.316	0.486	1.003	0.136	0.941	0.058	0.014	0.000	0.541	0.807	0.016
5	50	0.709	-0.519	0.423	1.482	-0.158	1.305	0.056	0.008	0.010	0.604	3.311	0.066
0	25	1.191	0.409	1.054	1.173	0.218	1.063	0.056	0.009	0.010	0.310	0.511	0.010
1	25	0.710	-0.180	0.563	2.771	0.823	2.541	0.057	0.008	0.009	0.354	27.920	0.558
2	25	0.635	-0.491	0.000	0.914	0.089	0.693	0.058	0.014	0.000	0.286	4.646	0.093
3	25	0.710	-0.180	0.563	2.788	0.832	2.555	0.058	0.014	0.000	0.352	26.850	0.537
4	25	0.710	-0.180	0.563	1.117	0.265	0.944	0.058	0.014	0.000	0.310	1.865	0.037
5	25	0.790	-0.626	0.377	1.570	-0.013	1.315	0.056	0.008	0.008	0.319	5.882	0.118

Appendix E
Item-Level Evaluation Criteria Summary Statistics

Table E.1. Item-Level Evaluation Criteria Summary Statistics, Condition 0, Sample Size = 1,000

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.265	0.121	0.025	0.118	-1.664	0.159	0.086	0.133	0.174	0.045	0.044	0.012	0.151
2	1.010	0.211	-0.080	0.195	-0.532	0.319	-0.242	0.208	0.249	0.094	-0.081	0.048	0.188
3	1.396	0.141	0.026	0.139	-1.816	0.153	0.064	0.139	0.180	0.042	0.040	0.013	0.156
4	1.292	0.135	0.022	0.133	-1.886	0.179	0.064	0.167	0.184	0.036	0.034	0.013	0.159
5	0.826	0.136	0.036	0.131	-2.842	0.377	0.098	0.364	0.193	0.015	0.013	0.006	0.164
6	1.275	0.145	0.015	0.145	-2.511	0.218	0.019	0.217	0.196	0.018	0.016	0.007	0.164
7	1.718	0.188	0.048	0.182	-1.310	0.130	0.070	0.109	0.175	0.039	0.035	0.018	0.167
8	2.853	0.338	0.133	0.310	-1.184	0.099	0.066	0.074	0.163	0.040	0.033	0.023	0.162
9	3.576	0.578	-0.074	0.573	-1.482	0.129	-0.032	0.125	0.260	0.070	-0.050	0.048	0.231
10	1.673	0.188	0.033	0.185	-2.161	0.163	0.050	0.155	0.194	0.027	0.024	0.012	0.171
11	1.925	0.202	0.075	0.188	-1.341	0.119	0.069	0.097	0.192	0.035	0.022	0.028	0.148
12	1.697	0.190	-0.013	0.189	-2.222	0.204	-0.092	0.182	0.217	0.075	-0.073	0.015	0.181
13	1.527	0.179	0.017	0.178	-1.952	0.167	0.029	0.165	0.202	0.019	0.012	0.015	0.182
14	1.722	0.297	-0.058	0.292	-1.078	0.178	-0.078	0.160	0.256	0.074	-0.054	0.051	0.175
15	0.788	0.106	0.018	0.104	-2.026	0.262	0.094	0.245	0.181	0.033	0.031	0.009	0.164
16	1.237	0.117	0.057	0.102	-0.834	0.140	0.096	0.101	0.179	0.037	0.029	0.024	0.127
17	2.019	0.264	0.039	0.261	-2.424	0.206	0.016	0.206	0.208	0.015	-0.012	0.010	0.211
18	1.855	0.181	0.055	0.172	-1.476	0.125	0.064	0.107	0.196	0.037	0.026	0.026	0.138
19	1.204	0.179	-0.066	0.167	-0.741	0.244	-0.161	0.184	0.249	0.078	-0.061	0.048	0.168
20	0.285	0.085	0.025	0.081	-3.698	1.090	0.122	1.083	0.201	0.003	0.001	0.003	0.175
21	1.349	0.137	0.069	0.118	-0.343	0.124	0.098	0.077	0.154	0.041	0.034	0.022	0.136
22	1.769	0.164	0.079	0.144	-1.125	0.133	0.106	0.081	0.163	0.048	0.043	0.022	0.155
23	1.436	0.212	0.046	0.206	0.276	0.120	0.016	0.119	0.231	0.038	-0.009	0.037	0.162
24	0.997	0.112	0.027	0.108	-1.811	0.195	0.059	0.186	0.188	0.030	0.028	0.010	0.163
25	2.320	0.313	0.100	0.297	-1.621	0.129	0.050	0.119	0.205	0.035	0.015	0.032	0.177
26	0.999	0.131	0.009	0.131	-0.881	0.167	-0.001	0.167	0.210	0.031	-0.010	0.029	0.138
27	1.692	0.191	0.032	0.188	-1.030	0.117	0.050	0.106	0.184	0.035	0.024	0.026	0.173
28	0.969	0.139	0.069	0.121	0.567	0.164	0.107	0.124	0.162	0.032	0.022	0.023	0.154
29	2.664	0.360	0.104	0.345	-1.891	0.132	0.069	0.113	0.181	0.044	0.041	0.016	0.202
30	1.627	0.163	0.087	0.138	-0.955	0.140	0.115	0.080	0.144	0.057	0.054	0.016	0.169
31	2.007	0.164	0.077	0.145	-0.643	0.094	0.077	0.054	0.109	0.051	0.049	0.014	0.166

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.950	0.263	-0.020	0.262	-0.263	0.119	-0.033	0.114	0.240	0.053	-0.030	0.044	0.188
33	1.779	0.185	0.099	0.157	-0.459	0.102	0.081	0.062	0.115	0.048	0.045	0.015	0.153
34	2.652	0.290	0.042	0.287	-1.710	0.115	0.040	0.108	0.194	0.036	0.024	0.026	0.175
35	2.540	0.298	0.080	0.287	-1.753	0.124	0.057	0.110	0.186	0.041	0.036	0.020	0.188
36	2.107	0.305	-0.023	0.304	1.464	0.106	0.024	0.103	0.159	0.018	-0.001	0.018	0.190
37	0.515	0.074	0.005	0.074	-0.778	0.221	0.002	0.221	0.205	0.012	-0.005	0.011	0.163
38	1.047	0.128	0.027	0.125	-1.013	0.162	0.047	0.155	0.202	0.027	0.012	0.024	0.149
39	1.009	0.293	0.049	0.289	1.589	0.231	-0.011	0.231	0.220	0.042	-0.010	0.041	0.229
40	1.309	0.214	-0.041	0.210	-0.367	0.184	-0.087	0.163	0.236	0.066	-0.044	0.049	0.179
41	0.851	0.106	0.001	0.106	-1.902	0.212	0.048	0.206	0.195	0.017	0.015	0.008	0.152
42	2.029	0.210	-0.012	0.210	-1.243	0.117	-0.003	0.117	0.214	0.039	-0.016	0.036	0.155
43	1.031	0.303	-0.119	0.279	0.435	0.294	-0.185	0.229	0.277	0.087	-0.063	0.061	0.221
44	0.594	0.090	-0.006	0.089	-2.075	0.329	-0.045	0.326	0.199	0.008	-0.002	0.007	0.165
45	2.783	0.413	-0.018	0.412	-1.450	0.128	-0.020	0.126	0.253	0.058	-0.038	0.045	0.198
46	1.783	0.172	0.083	0.150	-0.013	0.091	0.067	0.061	0.112	0.035	0.032	0.015	0.152
47	0.904	0.179	0.104	0.146	1.943	0.257	0.103	0.235	0.151	0.037	0.031	0.020	0.206
48	0.941	0.093	0.041	0.083	-0.434	0.161	0.136	0.087	0.160	0.052	0.050	0.017	0.131
49	0.837	0.099	0.047	0.088	-0.188	0.180	0.143	0.109	0.171	0.045	0.041	0.018	0.141
50	0.398	0.319	0.088	0.307	1.474	0.400	0.034	0.398	0.226	0.060	0.016	0.057	0.202

Table E.2. Item-Level Evaluation Criteria Summary Statistics, Condition 1, Sample Size = 1,000

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	0.990	0.280	-0.250	0.126	-2.092	0.405	-0.342	0.218	0.166	0.037	0.036	0.007	0.413
2	0.694	0.405	-0.396	0.085	-0.956	0.690	-0.666	0.178	0.186	0.146	-0.144	0.025	0.264
3	1.174	0.287	-0.196	0.209	-2.179	0.384	-0.299	0.241	0.178	0.040	0.038	0.012	0.392
4	0.986	0.318	-0.284	0.143	-2.420	0.547	-0.470	0.278	0.173	0.024	0.023	0.007	0.455
5	0.605	0.214	-0.185	0.108	-3.815	1.040	-0.875	0.563	0.188	0.010	0.008	0.005	0.286
6	0.991	0.328	-0.269	0.187	-3.189	0.787	-0.659	0.430	0.190	0.011	0.010	0.004	0.398
7	1.496	0.297	-0.174	0.241	-1.540	0.230	-0.160	0.165	0.174	0.039	0.034	0.020	0.358
8	1.999	0.722	-0.721	0.015	-1.524	0.282	-0.274	0.066	0.139	0.020	0.009	0.018	0.481
9	2.000	1.650	-1.650	0.000	-2.073	0.630	-0.623	0.096	0.172	0.139	-0.138	0.018	0.598
10	1.584	0.321	-0.056	0.316	-2.435	0.377	-0.225	0.302	0.201	0.035	0.031	0.016	0.349
11	1.743	0.254	-0.107	0.230	-1.533	0.182	-0.123	0.134	0.201	0.041	0.031	0.027	0.295
12	1.529	0.367	-0.181	0.319	-2.588	0.598	-0.458	0.385	0.217	0.076	-0.073	0.022	0.360
13	1.363	0.328	-0.147	0.294	-2.269	0.431	-0.289	0.320	0.208	0.035	0.018	0.030	0.385
14	1.270	0.550	-0.510	0.207	-1.480	0.517	-0.480	0.190	0.204	0.113	-0.106	0.037	0.303
15	0.631	0.163	-0.139	0.084	-2.518	0.504	-0.398	0.310	0.177	0.028	0.027	0.008	0.276
16	0.865	0.322	-0.315	0.068	-1.215	0.306	-0.285	0.112	0.144	0.013	-0.006	0.012	0.377
17	1.830	0.320	-0.150	0.283	-2.782	0.489	-0.342	0.349	0.206	0.017	-0.014	0.009	0.324
18	1.626	0.311	-0.174	0.257	-1.730	0.259	-0.190	0.176	0.198	0.038	0.028	0.025	0.333
19	0.807	0.471	-0.463	0.088	-1.243	0.683	-0.663	0.164	0.176	0.135	-0.134	0.020	0.298
20	0.266	0.056	0.006	0.056	-3.807	0.845	0.013	0.844	0.205	0.007	0.005	0.006	0.130
21	0.837	0.447	-0.443	0.060	-0.631	0.211	-0.191	0.089	0.113	0.012	-0.007	0.010	0.461
22	1.410	0.332	-0.280	0.177	-1.401	0.208	-0.171	0.119	0.150	0.032	0.030	0.013	0.362
23	0.846	0.549	-0.544	0.070	0.088	0.214	-0.172	0.127	0.154	0.089	-0.086	0.024	0.351
24	0.762	0.227	-0.208	0.091	-2.344	0.540	-0.474	0.258	0.176	0.018	0.016	0.007	0.358
25	1.991	0.237	-0.229	0.058	-1.894	0.243	-0.224	0.094	0.203	0.035	0.013	0.032	0.352
26	0.703	0.298	-0.288	0.078	-1.338	0.491	-0.458	0.177	0.165	0.056	-0.055	0.012	0.318
27	1.158	0.520	-0.502	0.133	-1.455	0.400	-0.375	0.140	0.145	0.019	-0.015	0.012	0.491
28	0.579	0.324	-0.322	0.043	0.611	0.214	0.151	0.152	0.118	0.025	-0.022	0.013	0.459
29	2.000	0.560	-0.560	0.000	-2.309	0.359	-0.349	0.087	0.168	0.033	0.028	0.016	0.556
30	1.229	0.336	-0.311	0.126	-1.238	0.197	-0.168	0.102	0.128	0.039	0.038	0.010	0.415
31	1.318	0.621	-0.612	0.105	-0.905	0.196	-0.185	0.065	0.093	0.034	0.033	0.008	0.523

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.258	0.720	-0.712	0.109	-0.535	0.325	-0.305	0.112	0.170	0.105	-0.100	0.031	0.291
33	1.121	0.567	-0.559	0.092	-0.704	0.180	-0.164	0.073	0.092	0.023	0.022	0.008	0.502
34	2.000	0.610	-0.610	0.000	-2.111	0.374	-0.361	0.095	0.172	0.020	0.002	0.020	0.548
35	2.000	0.460	-0.460	0.000	-2.122	0.324	-0.312	0.088	0.172	0.029	0.022	0.019	0.509
36	0.687	1.444	-1.443	0.064	2.458	1.055	1.018	0.277	0.105	0.057	-0.055	0.015	1.143
37	0.410	0.112	-0.101	0.048	-1.032	0.348	-0.252	0.240	0.192	0.020	-0.018	0.009	0.223
38	0.758	0.275	-0.262	0.082	-1.445	0.420	-0.385	0.168	0.166	0.026	-0.024	0.010	0.352
39	0.640	0.330	-0.321	0.078	1.900	0.399	0.300	0.263	0.182	0.059	-0.048	0.034	0.362
40	0.900	0.461	-0.450	0.097	-0.689	0.433	-0.409	0.143	0.172	0.111	-0.108	0.026	0.260
41	0.635	0.231	-0.215	0.084	-2.534	0.653	-0.584	0.292	0.181	0.006	0.001	0.006	0.343
42	1.680	0.436	-0.360	0.246	-1.516	0.321	-0.276	0.163	0.202	0.041	-0.028	0.030	0.340
43	0.690	0.470	-0.460	0.099	0.282	0.408	-0.338	0.228	0.217	0.131	-0.123	0.044	0.326
44	0.451	0.162	-0.149	0.063	-2.733	0.801	-0.703	0.384	0.188	0.013	-0.012	0.006	0.269
45	1.995	0.805	-0.805	0.026	-1.868	0.448	-0.438	0.090	0.201	0.094	-0.089	0.032	0.405
46	0.976	0.727	-0.724	0.063	-0.137	0.095	-0.057	0.077	0.079	0.008	-0.002	0.008	0.571
47	0.531	0.272	-0.269	0.040	2.656	0.860	0.816	0.272	0.119	0.017	-0.001	0.017	0.448
48	0.634	0.270	-0.266	0.045	-0.726	0.188	-0.156	0.104	0.132	0.024	0.022	0.010	0.421
49	0.614	0.182	-0.176	0.047	-0.357	0.121	-0.027	0.118	0.146	0.020	0.016	0.013	0.305
50	0.465	0.208	0.155	0.138	1.863	0.704	0.423	0.563	0.294	0.110	0.084	0.071	0.173

Table E.3. Item-Level Evaluation Criteria Summary Statistics, Condition 2, Sample Size = 1,000

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.000	0.240	-0.240	0.000	-1.956	0.230	-0.206	0.102	0.200	0.070	0.070	0.000	0.374
2	1.000	0.090	-0.090	0.000	-0.594	0.314	-0.304	0.081	0.200	0.130	-0.130	0.000	0.266
3	1.000	0.370	-0.370	0.000	-2.305	0.441	-0.425	0.118	0.200	0.060	0.060	0.000	0.510
4	1.000	0.270	-0.270	0.000	-2.277	0.346	-0.327	0.114	0.200	0.050	0.050	0.000	0.390
5	1.000	0.210	0.210	0.000	-2.447	0.506	0.493	0.113	0.200	0.020	0.020	0.000	0.233
6	1.000	0.260	-0.260	0.000	-3.051	0.536	-0.521	0.129	0.200	0.020	0.020	0.000	0.334
7	1.000	0.670	-0.670	0.000	-1.858	0.489	-0.478	0.102	0.200	0.060	0.060	0.000	0.738
8	1.000	1.720	-1.720	0.000	-2.073	0.828	-0.823	0.097	0.200	0.070	0.070	0.000	1.260
9	1.000	2.650	-2.650	0.000	-2.919	1.475	-1.469	0.141	0.200	0.110	-0.110	0.000	1.179
10	1.000	0.640	-0.640	0.000	-3.088	0.887	-0.878	0.127	0.200	0.030	0.030	0.000	0.665
11	1.000	0.850	-0.850	0.000	-2.057	0.655	-0.647	0.103	0.200	0.030	0.030	0.000	0.812
12	1.000	0.710	-0.710	0.000	-3.244	1.123	-1.114	0.137	0.200	0.090	-0.090	0.000	0.606
13	1.000	0.510	-0.510	0.000	-2.657	0.686	-0.677	0.115	0.200	0.010	0.010	0.000	0.569
14	1.000	0.780	-0.780	0.000	-1.672	0.678	-0.672	0.089	0.200	0.110	-0.110	0.000	0.430
15	1.000	0.230	0.230	0.000	-1.636	0.494	0.484	0.098	0.200	0.050	0.050	0.000	0.271
16	1.000	0.180	-0.180	0.000	-0.922	0.102	0.008	0.101	0.200	0.050	0.050	0.000	0.276
17	1.000	0.980	-0.980	0.000	-3.826	1.395	-1.386	0.159	0.200	0.020	-0.020	0.000	0.727
18	1.000	0.800	-0.800	0.000	-2.228	0.696	-0.688	0.105	0.200	0.030	0.030	0.000	0.791
19	1.000	0.270	-0.270	0.000	-0.936	0.367	-0.356	0.089	0.200	0.110	-0.110	0.000	0.148
20	1.000	0.740	0.740	0.000	-1.021	2.800	2.799	0.097	0.200	0.000	0.000	0.000	1.166
21	1.000	0.280	-0.280	0.000	-0.296	0.167	0.144	0.084	0.200	0.080	0.080	0.000	0.419
22	1.000	0.690	-0.690	0.000	-1.596	0.377	-0.366	0.088	0.200	0.080	0.080	0.000	0.751
23	1.000	0.390	-0.390	0.000	0.267	0.084	0.007	0.084	0.200	0.040	-0.040	0.000	0.265
24	1.000	0.030	0.030	0.000	-1.799	0.122	0.071	0.099	0.200	0.040	0.040	0.000	0.089
25	1.000	1.220	-1.220	0.000	-2.708	1.045	-1.038	0.119	0.200	0.010	0.010	0.000	1.024
26	1.000	0.010	0.010	0.000	-0.863	0.095	0.017	0.094	0.200	0.020	-0.020	0.000	0.113
27	1.000	0.660	-0.660	0.000	-1.446	0.378	-0.366	0.093	0.200	0.040	0.040	0.000	0.652
28	1.000	0.100	0.100	0.000	0.790	0.345	0.330	0.101	0.200	0.060	0.060	0.000	0.181
29	1.000	1.560	-1.560	0.000	-3.258	1.305	-1.298	0.137	0.200	0.060	0.060	0.000	1.213
30	1.000	0.540	-0.540	0.000	-1.239	0.193	-0.169	0.092	0.200	0.110	0.110	0.000	0.696
31	1.000	0.930	-0.930	0.000	-0.844	0.153	-0.124	0.090	0.200	0.140	0.140	0.000	0.983

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.000	0.970	-0.970	0.000	-0.539	0.321	-0.309	0.086	0.200	0.070	-0.070	0.000	0.436
33	1.000	0.680	-0.680	0.000	-0.504	0.091	0.036	0.084	0.200	0.130	0.130	0.000	0.811
34	1.000	1.610	-1.610	0.000	-2.972	1.231	-1.222	0.145	0.200	0.030	0.030	0.000	1.228
35	1.000	1.460	-1.460	0.000	-2.989	1.186	-1.179	0.134	0.200	0.050	0.050	0.000	1.198
36	1.000	1.130	-1.130	0.000	2.499	1.069	1.059	0.140	0.200	0.040	0.040	0.000	1.063
37	1.000	0.490	0.490	0.000	-0.312	0.479	0.468	0.101	0.200	0.010	-0.010	0.000	0.696
38	1.000	0.020	-0.020	0.000	-1.027	0.088	0.034	0.081	0.200	0.010	0.010	0.000	0.081
39	1.000	0.040	0.040	0.000	1.534	0.128	-0.066	0.109	0.200	0.030	-0.030	0.000	0.178
40	1.000	0.350	-0.350	0.000	-0.515	0.249	-0.235	0.084	0.200	0.080	-0.080	0.000	0.178
41	1.000	0.150	0.150	0.000	-1.661	0.305	0.289	0.095	0.200	0.020	0.020	0.000	0.198
42	1.000	1.040	-1.040	0.000	-1.994	0.761	-0.754	0.109	0.200	0.030	-0.030	0.000	0.780
43	1.000	0.150	-0.150	0.000	0.263	0.367	-0.357	0.085	0.200	0.140	-0.140	0.000	0.389
44	1.000	0.400	0.400	0.000	-1.280	0.757	0.750	0.099	0.200	0.000	0.000	0.000	0.564
45	1.000	1.800	-1.800	0.000	-2.673	1.249	-1.243	0.124	0.200	0.090	-0.090	0.000	1.021
46	1.000	0.700	-0.700	0.000	0.193	0.287	0.273	0.087	0.200	0.120	0.120	0.000	0.795
47	1.000	0.200	0.200	0.000	2.255	0.435	0.415	0.130	0.200	0.080	0.080	0.000	0.272
48	1.000	0.100	0.100	0.000	-0.251	0.329	0.319	0.079	0.200	0.090	0.090	0.000	0.134
49	1.000	0.210	0.210	0.000	0.004	0.346	0.334	0.090	0.200	0.070	0.070	0.000	0.168
50	1.000	0.690	0.690	0.000	0.739	0.709	-0.701	0.105	0.200	0.010	-0.010	0.000	0.986

Table E.4. Item-Level Evaluation Criteria Summary Statistics, Condition 3, Sample Size = 1,000

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	0.990	0.280	-0.250	0.126	-2.042	0.367	-0.292	0.221	0.200	0.070	0.070	0.000	0.442
2	0.694	0.405	-0.396	0.085	-0.914	0.640	-0.624	0.140	0.200	0.130	-0.130	0.000	0.272
3	1.174	0.287	-0.196	0.209	-2.155	0.363	-0.275	0.237	0.200	0.060	0.060	0.000	0.414
4	0.986	0.318	-0.284	0.143	-2.383	0.516	-0.433	0.280	0.200	0.050	0.050	0.000	0.472
5	0.605	0.214	-0.185	0.108	-3.791	1.023	-0.851	0.568	0.200	0.020	0.020	0.000	0.289
6	0.991	0.328	-0.269	0.187	-3.183	0.781	-0.653	0.428	0.200	0.020	0.020	0.000	0.405
7	1.496	0.297	-0.174	0.241	-1.513	0.204	-0.133	0.154	0.200	0.060	0.060	0.000	0.396
8	1.999	0.722	-0.721	0.015	-1.475	0.235	-0.225	0.068	0.200	0.070	0.070	0.000	0.576
9	2.000	1.650	-1.650	0.000	-2.060	0.618	-0.610	0.098	0.200	0.110	-0.110	0.000	0.620
10	1.584	0.321	-0.056	0.316	-2.447	0.377	-0.237	0.293	0.200	0.030	0.030	0.000	0.358
11	1.743	0.254	-0.107	0.230	-1.542	0.180	-0.132	0.122	0.200	0.030	0.030	0.000	0.303
12	1.529	0.367	-0.181	0.319	-2.616	0.609	-0.486	0.367	0.200	0.090	-0.090	0.000	0.366
13	1.363	0.328	-0.147	0.294	-2.286	0.424	-0.306	0.293	0.200	0.010	0.010	0.000	0.397
14	1.270	0.550	-0.510	0.207	-1.491	0.515	-0.491	0.155	0.200	0.110	-0.110	0.000	0.314
15	0.631	0.163	-0.139	0.084	-2.462	0.469	-0.342	0.321	0.200	0.050	0.050	0.000	0.286
16	0.865	0.322	-0.315	0.068	-1.096	0.204	-0.166	0.120	0.200	0.050	0.050	0.000	0.451
17	1.830	0.320	-0.150	0.283	-2.799	0.497	-0.359	0.343	0.200	0.020	-0.020	0.000	0.332
18	1.626	0.311	-0.174	0.257	-1.735	0.251	-0.195	0.159	0.200	0.030	0.030	0.000	0.345
19	0.807	0.471	-0.463	0.088	-1.189	0.626	-0.609	0.145	0.200	0.110	-0.110	0.000	0.315
20	0.266	0.056	0.006	0.056	-3.846	0.865	-0.026	0.864	0.200	0.000	0.000	0.000	0.129
21	0.837	0.447	-0.443	0.060	-0.415	0.101	0.025	0.098	0.200	0.080	0.080	0.000	0.605
22	1.410	0.332	-0.280	0.177	-1.339	0.159	-0.109	0.116	0.200	0.080	0.080	0.000	0.437
23	0.846	0.549	-0.544	0.070	0.235	0.101	-0.025	0.098	0.200	0.040	-0.040	0.000	0.391
24	0.762	0.227	-0.208	0.091	-2.297	0.503	-0.427	0.265	0.200	0.040	0.040	0.000	0.373
25	1.991	0.237	-0.229	0.058	-1.909	0.255	-0.239	0.089	0.200	0.010	0.010	0.000	0.356
26	0.703	0.298	-0.288	0.078	-1.246	0.406	-0.366	0.176	0.200	0.020	-0.020	0.000	0.352
27	1.158	0.520	-0.502	0.133	-1.374	0.326	-0.294	0.141	0.200	0.040	0.040	0.000	0.559
28	0.579	0.324	-0.322	0.043	1.027	0.591	0.567	0.166	0.200	0.060	0.060	0.000	0.598
29	2.000	0.560	-0.560	0.000	-2.295	0.348	-0.335	0.091	0.200	0.060	0.060	0.000	0.595
30	1.229	0.336	-0.311	0.126	-1.132	0.122	-0.062	0.105	0.200	0.110	0.110	0.000	0.531
31	1.318	0.621	-0.612	0.105	-0.751	0.078	-0.031	0.071	0.200	0.140	0.140	0.000	0.744

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.258	0.720	-0.712	0.109	-0.482	0.267	-0.252	0.087	0.200	0.070	-0.070	0.000	0.286
33	1.121	0.567	-0.559	0.092	-0.512	0.083	0.028	0.078	0.200	0.130	0.130	0.000	0.725
34	2.000	0.610	-0.610	0.000	-2.098	0.361	-0.348	0.097	0.200	0.030	0.030	0.000	0.581
35	2.000	0.460	-0.460	0.000	-2.109	0.313	-0.299	0.092	0.200	0.050	0.050	0.000	0.544
36	0.687	1.444	-1.443	0.064	3.335	1.927	1.895	0.350	0.200	0.040	0.040	0.000	1.390
37	0.410	0.112	-0.101	0.048	-0.998	0.333	-0.218	0.252	0.200	0.010	-0.010	0.000	0.224
38	0.758	0.275	-0.262	0.082	-1.364	0.347	-0.304	0.167	0.200	0.010	0.010	0.000	0.382
39	0.640	0.330	-0.321	0.078	2.030	0.479	0.430	0.210	0.200	0.030	-0.030	0.000	0.386
40	0.900	0.461	-0.450	0.097	-0.622	0.360	-0.342	0.112	0.200	0.080	-0.080	0.000	0.275
41	0.635	0.231	-0.215	0.084	-2.489	0.615	-0.539	0.297	0.200	0.020	0.020	0.000	0.352
42	1.680	0.436	-0.360	0.246	-1.525	0.319	-0.285	0.144	0.200	0.030	-0.030	0.000	0.350
43	0.690	0.470	-0.460	0.099	0.214	0.421	-0.406	0.111	0.200	0.140	-0.140	0.000	0.335
44	0.451	0.162	-0.149	0.063	-2.693	0.773	-0.663	0.398	0.200	0.000	0.000	0.000	0.273
45	1.995	0.805	-0.805	0.026	-1.882	0.461	-0.452	0.086	0.200	0.090	-0.090	0.000	0.415
46	0.976	0.727	-0.724	0.063	0.151	0.249	0.231	0.092	0.200	0.120	0.120	0.000	0.829
47	0.531	0.272	-0.269	0.040	3.565	1.753	1.725	0.315	0.200	0.080	0.080	0.000	0.676
48	0.634	0.270	-0.266	0.045	-0.499	0.138	0.071	0.118	0.200	0.090	0.090	0.000	0.516
49	0.614	0.182	-0.176	0.047	-0.151	0.222	0.179	0.132	0.200	0.070	0.070	0.000	0.381
50	0.465	0.208	0.155	0.138	1.042	0.455	-0.399	0.220	0.200	0.010	-0.010	0.000	0.336

Table E.5. Item-Level Evaluation Criteria Summary Statistics, Condition 4, Sample Size = 1,000

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	0.990	0.280	-0.250	0.126	-1.451	0.322	0.299	0.118	0.200	0.070	0.070	0.000	0.333
2	0.694	0.405	-0.396	0.085	-0.489	0.254	-0.199	0.158	0.200	0.130	-0.130	0.000	0.395
3	1.174	0.287	-0.196	0.209	-1.540	0.363	0.340	0.128	0.200	0.060	0.060	0.000	0.344
4	0.986	0.318	-0.284	0.143	-1.648	0.337	0.302	0.149	0.200	0.050	0.050	0.000	0.362
5	0.605	0.214	-0.185	0.108	-2.414	0.608	0.526	0.306	0.200	0.020	0.020	0.000	0.502
6	0.991	0.328	-0.269	0.188	-2.132	0.453	0.398	0.215	0.200	0.020	0.020	0.000	0.467
7	1.496	0.297	-0.175	0.241	-1.158	0.239	0.222	0.089	0.200	0.060	0.060	0.000	0.282
8	1.999	0.722	-0.721	0.015	-1.058	0.198	0.192	0.048	0.200	0.070	0.070	0.000	0.378
9	2.000	1.650	-1.650	0.000	-1.353	0.122	0.097	0.074	0.200	0.110	-0.110	0.000	0.435
10	1.584	0.321	-0.056	0.316	-1.755	0.488	0.455	0.177	0.200	0.030	0.030	0.000	0.522
11	1.743	0.254	-0.107	0.230	-1.166	0.260	0.244	0.091	0.200	0.030	0.030	0.000	0.289
12	1.530	0.366	-0.181	0.319	-1.833	0.366	0.297	0.214	0.200	0.090	-0.090	0.000	0.548
13	1.363	0.328	-0.147	0.293	-1.628	0.386	0.352	0.158	0.200	0.010	0.010	0.000	0.439
14	1.271	0.550	-0.509	0.207	-1.022	0.112	-0.022	0.110	0.200	0.110	-0.110	0.000	0.347
15	0.630	0.163	-0.140	0.084	-1.648	0.503	0.472	0.173	0.200	0.050	0.050	0.000	0.378
16	0.866	0.322	-0.314	0.068	-0.810	0.146	0.120	0.082	0.200	0.050	0.050	0.000	0.373
17	1.830	0.321	-0.150	0.283	-1.922	0.574	0.518	0.246	0.200	0.020	-0.020	0.000	0.726
18	1.626	0.311	-0.175	0.257	-1.292	0.269	0.248	0.104	0.200	0.030	0.030	0.000	0.309
19	0.807	0.471	-0.463	0.088	-0.723	0.200	-0.143	0.140	0.200	0.110	-0.110	0.000	0.361
20	0.266	0.056	0.006	0.056	-2.324	1.576	1.496	0.495	0.200	0.000	0.000	0.000	0.478
21	0.837	0.447	-0.443	0.060	-0.396	0.084	0.044	0.071	0.200	0.080	0.080	0.000	0.599
22	1.410	0.332	-0.281	0.177	-1.045	0.197	0.185	0.066	0.200	0.080	0.080	0.000	0.320
23	0.846	0.548	-0.544	0.070	0.215	0.129	-0.045	0.121	0.200	0.040	-0.040	0.000	0.394
24	0.762	0.227	-0.208	0.091	-1.552	0.349	0.318	0.145	0.200	0.040	0.040	0.000	0.344
25	1.991	0.236	-0.229	0.058	-1.319	0.359	0.351	0.077	0.200	0.010	0.010	0.000	0.452
26	0.703	0.298	-0.287	0.079	-0.806	0.152	0.074	0.133	0.200	0.020	-0.020	0.000	0.353
27	1.157	0.520	-0.503	0.133	-1.011	0.108	0.070	0.083	0.200	0.040	0.040	0.000	0.398
28	0.578	0.324	-0.322	0.043	0.522	0.154	0.062	0.142	0.200	0.060	0.060	0.000	0.609
29	2.000	0.560	-0.560	0.000	-1.558	0.406	0.402	0.060	0.200	0.060	0.060	0.000	0.503
30	1.229	0.336	-0.311	0.126	-0.909	0.173	0.162	0.063	0.200	0.110	0.110	0.000	0.431
31	1.318	0.621	-0.612	0.104	-0.686	0.058	0.034	0.047	0.200	0.140	0.140	0.000	0.710

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.257	0.721	-0.713	0.109	-0.330	0.136	-0.100	0.093	0.200	0.070	-0.070	0.000	0.291
33	1.121	0.566	-0.559	0.092	-0.517	0.060	0.023	0.055	0.200	0.130	0.130	0.000	0.725
34	2.000	0.610	-0.610	0.000	-1.428	0.328	0.322	0.064	0.200	0.030	0.030	0.000	0.436
35	2.000	0.460	-0.460	0.000	-1.441	0.374	0.369	0.059	0.200	0.050	0.050	0.000	0.453
36	0.687	1.444	-1.443	0.064	1.948	0.551	0.508	0.213	0.200	0.040	0.040	0.000	1.000
37	0.409	0.111	-0.101	0.048	-0.544	0.295	0.236	0.178	0.200	0.010	-0.010	0.000	0.283
38	0.757	0.275	-0.263	0.081	-0.923	0.181	0.137	0.118	0.200	0.010	0.010	0.000	0.352
39	0.640	0.330	-0.320	0.078	1.483	0.231	-0.117	0.199	0.200	0.030	-0.030	0.000	0.323
40	0.900	0.461	-0.451	0.097	-0.391	0.167	-0.111	0.125	0.200	0.080	-0.080	0.000	0.321
41	0.634	0.232	-0.216	0.084	-1.633	0.357	0.317	0.162	0.200	0.020	0.020	0.000	0.398
42	1.680	0.436	-0.360	0.246	-1.142	0.139	0.098	0.098	0.200	0.030	-0.030	0.000	0.277
43	0.690	0.470	-0.460	0.099	0.385	0.306	-0.235	0.196	0.200	0.140	-0.140	0.000	0.425
44	0.451	0.162	-0.149	0.063	-1.703	0.398	0.327	0.227	0.200	0.000	0.000	0.000	0.407
45	1.995	0.805	-0.805	0.025	-1.261	0.186	0.169	0.078	0.200	0.090	-0.090	0.000	0.438
46	0.976	0.727	-0.724	0.063	-0.087	0.063	-0.007	0.063	0.200	0.120	0.120	0.000	0.881
47	0.532	0.271	-0.268	0.040	1.924	0.222	0.084	0.206	0.200	0.080	0.080	0.000	0.731
48	0.634	0.270	-0.267	0.044	-0.448	0.144	0.122	0.077	0.200	0.090	0.090	0.000	0.506
49	0.614	0.182	-0.176	0.046	-0.201	0.161	0.130	0.096	0.200	0.070	0.070	0.000	0.385
50	0.465	0.208	0.155	0.138	1.389	0.372	-0.051	0.368	0.200	0.010	-0.010	0.000	0.427

Table E.6. Item-Level Evaluation Criteria Summary Statistics, Condition 5, Sample Size = 1,000

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.221	0.116	-0.019	0.114	-1.702	0.146	0.048	0.138	0.174	0.045	0.044	0.012	0.162
2	0.974	0.216	-0.116	0.183	-0.550	0.332	-0.260	0.207	0.244	0.097	-0.086	0.046	0.186
3	1.343	0.134	-0.027	0.132	-1.863	0.145	0.017	0.144	0.180	0.041	0.040	0.012	0.171
4	1.245	0.129	-0.025	0.127	-1.932	0.175	0.018	0.174	0.184	0.036	0.034	0.013	0.172
5	0.796	0.127	0.006	0.126	-2.922	0.383	0.018	0.382	0.194	0.015	0.014	0.006	0.157
6	1.220	0.142	-0.041	0.136	-2.588	0.234	-0.058	0.227	0.197	0.018	0.017	0.007	0.170
7	1.646	0.171	-0.024	0.170	-1.343	0.119	0.037	0.114	0.173	0.037	0.033	0.018	0.181
8	2.690	0.284	-0.030	0.282	-1.215	0.085	0.035	0.077	0.160	0.037	0.030	0.022	0.159
9	3.191	0.635	-0.459	0.439	-1.552	0.165	-0.102	0.129	0.245	0.078	-0.065	0.044	0.231
10	1.582	0.177	-0.058	0.167	-2.233	0.163	-0.023	0.162	0.193	0.025	0.023	0.011	0.180
11	1.832	0.173	-0.018	0.172	-1.380	0.104	0.030	0.100	0.188	0.032	0.018	0.026	0.150
12	1.598	0.200	-0.113	0.166	-2.305	0.256	-0.175	0.186	0.214	0.077	-0.076	0.012	0.176
13	1.458	0.171	-0.052	0.163	-2.011	0.171	-0.031	0.168	0.200	0.016	0.010	0.013	0.190
14	1.643	0.294	-0.137	0.260	-1.111	0.195	-0.111	0.160	0.251	0.077	-0.059	0.049	0.169
15	0.765	0.101	-0.005	0.101	-2.074	0.258	0.046	0.254	0.181	0.032	0.031	0.009	0.163
16	1.201	0.101	0.021	0.099	-0.846	0.134	0.084	0.104	0.178	0.036	0.028	0.023	0.130
17	1.856	0.260	-0.125	0.228	-2.531	0.237	-0.091	0.218	0.207	0.015	-0.013	0.008	0.196
18	1.767	0.160	-0.033	0.156	-1.519	0.112	0.021	0.110	0.193	0.033	0.023	0.024	0.145
19	1.161	0.191	-0.109	0.157	-0.760	0.257	-0.180	0.184	0.245	0.079	-0.065	0.046	0.163
20	0.274	0.079	0.014	0.078	-3.839	1.147	-0.019	1.147	0.201	0.004	0.001	0.004	0.166
21	1.315	0.121	0.035	0.115	-0.340	0.127	0.100	0.079	0.153	0.039	0.033	0.021	0.141
22	1.702	0.136	0.012	0.136	-1.148	0.117	0.082	0.083	0.160	0.045	0.040	0.021	0.157
23	1.386	0.195	-0.004	0.195	0.285	0.122	0.025	0.120	0.226	0.039	-0.014	0.037	0.166
24	0.964	0.105	-0.006	0.104	-1.853	0.193	0.017	0.193	0.188	0.030	0.028	0.010	0.164
25	2.171	0.269	-0.049	0.265	-1.679	0.123	-0.009	0.123	0.200	0.030	0.010	0.028	0.180
26	0.968	0.128	-0.023	0.126	-0.900	0.171	-0.020	0.170	0.207	0.031	-0.013	0.028	0.137
27	1.634	0.181	-0.026	0.179	-1.050	0.114	0.030	0.110	0.182	0.034	0.022	0.026	0.182
28	0.947	0.127	0.047	0.117	0.587	0.178	0.127	0.124	0.161	0.031	0.021	0.023	0.160
29	2.467	0.318	-0.093	0.304	-1.957	0.122	0.003	0.122	0.178	0.040	0.038	0.015	0.217
30	1.575	0.137	0.035	0.132	-0.970	0.131	0.100	0.083	0.143	0.055	0.053	0.016	0.180
31	1.946	0.140	0.016	0.139	-0.646	0.093	0.074	0.057	0.109	0.051	0.049	0.014	0.175

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.870	0.261	-0.101	0.241	-0.267	0.119	-0.037	0.113	0.235	0.055	-0.035	0.042	0.186
33	1.730	0.159	0.050	0.151	-0.458	0.104	0.083	0.064	0.115	0.047	0.045	0.015	0.159
34	2.461	0.294	-0.149	0.253	-1.772	0.116	-0.022	0.113	0.188	0.028	0.018	0.022	0.205
35	2.369	0.274	-0.091	0.259	-1.813	0.118	-0.003	0.118	0.181	0.036	0.031	0.018	0.211
36	2.012	0.302	-0.118	0.278	1.502	0.123	0.062	0.107	0.156	0.017	-0.004	0.017	0.209
37	0.499	0.073	-0.011	0.072	-0.800	0.228	-0.020	0.228	0.203	0.013	-0.008	0.011	0.166
38	1.016	0.119	-0.004	0.119	-1.032	0.160	0.028	0.157	0.200	0.025	0.010	0.023	0.152
39	0.915	0.253	-0.045	0.249	1.638	0.249	0.038	0.246	0.209	0.043	-0.021	0.037	0.238
40	1.255	0.212	-0.095	0.190	-0.381	0.188	-0.101	0.158	0.230	0.068	-0.050	0.046	0.177
41	0.825	0.106	-0.025	0.103	-1.946	0.215	0.004	0.215	0.194	0.017	0.014	0.008	0.156
42	1.924	0.220	-0.116	0.187	-1.282	0.126	-0.042	0.119	0.208	0.040	-0.022	0.033	0.157
43	0.948	0.315	-0.202	0.241	0.407	0.298	-0.213	0.208	0.261	0.095	-0.079	0.052	0.238
44	0.576	0.091	-0.025	0.088	-2.127	0.352	-0.097	0.339	0.198	0.008	-0.002	0.007	0.166
45	2.560	0.428	-0.240	0.354	-1.514	0.156	-0.084	0.132	0.240	0.064	-0.050	0.041	0.194
46	1.743	0.153	0.043	0.147	0.001	0.101	0.081	0.061	0.110	0.034	0.030	0.014	0.157
47	0.861	0.151	0.061	0.138	1.998	0.292	0.158	0.246	0.147	0.032	0.027	0.018	0.203
48	0.919	0.084	0.019	0.082	-0.435	0.162	0.135	0.089	0.159	0.052	0.049	0.017	0.138
49	0.815	0.089	0.025	0.086	-0.186	0.182	0.144	0.111	0.169	0.043	0.039	0.018	0.144
50	0.319	0.127	0.009	0.127	1.490	0.429	0.050	0.426	0.208	0.029	-0.002	0.029	0.184

Table E.7. Item-Level Evaluation Criteria Summary Statistics, Condition 0, Sample Size = 200

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.277	0.346	0.037	0.344	-1.716	0.484	0.034	0.482	0.194	0.064	0.064	0.008	0.366
2	0.939	0.361	-0.151	0.328	-0.703	0.532	-0.413	0.336	0.208	0.123	-0.122	0.019	0.389
3	1.402	0.406	0.032	0.404	-1.887	0.412	-0.007	0.412	0.196	0.057	0.056	0.009	0.401
4	1.293	0.405	0.023	0.404	-1.981	0.609	-0.031	0.608	0.196	0.047	0.046	0.008	0.385
5	0.831	0.303	0.041	0.300	-3.055	1.068	-0.115	1.061	0.198	0.019	0.018	0.005	0.374
6	1.257	0.479	-0.003	0.479	-2.830	1.058	-0.300	1.014	0.199	0.020	0.019	0.006	0.487
7	1.818	0.560	0.148	0.540	-1.284	0.265	0.097	0.247	0.196	0.058	0.056	0.016	0.312
8	3.022	0.967	0.302	0.919	-1.135	0.238	0.115	0.209	0.189	0.064	0.059	0.024	0.401
9	3.274	0.926	-0.376	0.846	-1.587	0.338	-0.137	0.309	0.209	0.105	-0.102	0.029	0.444
10	1.688	0.518	0.048	0.516	-2.209	0.460	0.001	0.460	0.198	0.030	0.028	0.008	0.410
11	2.060	0.655	0.210	0.620	-1.337	0.281	0.073	0.271	0.198	0.034	0.028	0.019	0.405
12	1.719	0.554	0.009	0.554	-2.382	0.694	-0.252	0.647	0.204	0.086	-0.086	0.009	0.493
13	1.574	0.437	0.064	0.433	-1.981	0.404	-0.001	0.404	0.199	0.013	0.009	0.009	0.388
14	1.584	0.452	-0.196	0.407	-1.207	0.338	-0.207	0.267	0.206	0.105	-0.104	0.019	0.354
15	0.775	0.247	0.005	0.247	-2.180	0.695	-0.060	0.693	0.197	0.047	0.047	0.007	0.336
16	1.252	0.299	0.072	0.290	-0.801	0.257	0.129	0.222	0.191	0.044	0.041	0.015	0.280
17	1.957	0.579	-0.023	0.578	-2.568	0.617	-0.128	0.603	0.202	0.019	-0.019	0.006	0.454
18	1.945	0.508	0.145	0.487	-1.467	0.290	0.073	0.280	0.201	0.036	0.031	0.018	0.365
19	1.078	0.343	-0.192	0.285	-0.927	0.476	-0.347	0.325	0.206	0.106	-0.105	0.015	0.354
20	0.330	0.148	0.070	0.130	-3.414	1.599	0.406	1.546	0.201	0.002	0.001	0.002	0.314
21	1.416	0.339	0.136	0.310	-0.240	0.258	0.200	0.164	0.184	0.068	0.064	0.024	0.303
22	1.797	0.447	0.107	0.434	-1.082	0.304	0.148	0.265	0.189	0.071	0.069	0.018	0.385
23	1.307	0.486	-0.083	0.479	0.257	0.215	-0.003	0.215	0.204	0.047	-0.036	0.030	0.352
24	0.988	0.283	0.018	0.282	-1.864	0.522	0.007	0.522	0.197	0.038	0.037	0.008	0.338
25	2.359	0.719	0.139	0.705	-1.668	0.285	0.002	0.285	0.203	0.021	0.013	0.017	0.400
26	1.017	0.237	0.027	0.235	-0.908	0.269	-0.028	0.268	0.203	0.023	-0.017	0.016	0.289
27	1.761	0.473	0.101	0.462	-0.992	0.255	0.088	0.239	0.196	0.040	0.036	0.019	0.364
28	0.997	0.314	0.097	0.299	0.746	0.426	0.286	0.317	0.179	0.044	0.039	0.020	0.392
29	2.635	0.615	0.075	0.611	-1.925	0.330	0.035	0.328	0.195	0.057	0.055	0.012	0.396
30	1.702	0.441	0.162	0.410	-0.884	0.270	0.187	0.196	0.182	0.094	0.092	0.017	0.369
31	2.171	0.600	0.241	0.549	-0.545	0.227	0.175	0.145	0.157	0.099	0.097	0.019	0.398

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.895	0.530	-0.075	0.524	-0.293	0.199	-0.063	0.189	0.216	0.067	-0.054	0.040	0.300
33	1.827	0.419	0.147	0.392	-0.346	0.247	0.194	0.154	0.161	0.093	0.091	0.018	0.377
34	2.603	0.631	-0.007	0.631	-1.727	0.232	0.023	0.230	0.199	0.032	0.029	0.014	0.341
35	2.563	0.628	0.103	0.620	-1.784	0.267	0.026	0.266	0.194	0.046	0.044	0.014	0.403
36	1.780	0.610	-0.350	0.499	1.708	0.465	0.268	0.381	0.159	0.026	-0.001	0.026	0.561
37	0.532	0.173	0.022	0.172	-0.750	0.432	0.031	0.431	0.201	0.012	-0.009	0.008	0.330
38	1.033	0.321	0.013	0.320	-1.067	0.353	-0.007	0.353	0.202	0.024	0.012	0.020	0.375
39	0.885	0.408	-0.076	0.401	1.806	0.797	0.206	0.770	0.201	0.041	-0.029	0.029	0.452
40	1.206	0.315	-0.144	0.280	-0.429	0.286	-0.149	0.244	0.204	0.079	-0.076	0.021	0.338
41	0.829	0.243	-0.021	0.242	-2.112	0.748	-0.162	0.731	0.199	0.020	0.019	0.005	0.340
42	2.021	0.615	-0.020	0.615	-1.305	0.298	-0.065	0.291	0.204	0.034	-0.026	0.023	0.388
43	0.897	0.397	-0.253	0.306	0.317	0.412	-0.304	0.279	0.212	0.131	-0.129	0.025	0.453
44	0.583	0.217	-0.017	0.216	-2.346	1.085	-0.316	1.038	0.200	0.004	0.000	0.004	0.406
45	2.692	0.747	-0.108	0.739	-1.574	0.338	-0.144	0.306	0.210	0.083	-0.080	0.022	0.401
46	1.944	0.478	0.244	0.411	0.115	0.246	0.195	0.151	0.153	0.075	0.073	0.019	0.367
47	0.966	0.380	0.166	0.341	2.357	1.118	0.517	0.992	0.173	0.058	0.053	0.024	0.484
48	0.921	0.231	0.021	0.230	-0.341	0.348	0.229	0.262	0.190	0.081	0.080	0.012	0.342
49	0.897	0.245	0.107	0.220	-0.088	0.322	0.242	0.212	0.189	0.062	0.059	0.016	0.293
50	0.405	0.217	0.095	0.195	1.344	0.775	-0.096	0.769	0.209	0.014	-0.002	0.014	0.384

Table E.8. Item-Level Evaluation Criteria Summary Statistics, Condition 1, Sample Size = 200

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.108	0.431	-0.132	0.410	-2.093	0.856	-0.343	0.784	0.194	0.064	0.064	0.007	0.507
2	0.662	0.474	-0.428	0.203	-1.026	0.843	-0.736	0.411	0.192	0.138	-0.138	0.009	0.433
3	1.245	0.494	-0.125	0.478	-2.268	0.795	-0.388	0.694	0.198	0.059	0.058	0.008	0.542
4	1.073	0.446	-0.197	0.400	-2.441	0.986	-0.491	0.855	0.195	0.045	0.045	0.007	0.527
5	0.678	0.346	-0.112	0.328	-3.990	2.028	-1.050	1.735	0.198	0.019	0.018	0.004	0.381
6	1.058	0.547	-0.202	0.509	-3.220	2.764	-0.690	2.676	0.200	0.020	0.020	0.005	0.907
7	1.526	0.446	-0.144	0.422	-1.534	0.370	-0.154	0.337	0.192	0.054	0.052	0.013	0.430
8	1.935	0.809	-0.785	0.194	-1.457	0.293	-0.207	0.207	0.178	0.051	0.048	0.019	0.574
9	1.952	1.707	-1.698	0.173	-2.057	0.675	-0.607	0.296	0.191	0.120	-0.119	0.015	0.625
10	1.497	0.498	-0.143	0.477	-2.600	0.847	-0.390	0.752	0.199	0.030	0.029	0.008	0.446
11	1.682	0.427	-0.168	0.393	-1.600	0.398	-0.190	0.350	0.194	0.029	0.024	0.017	0.485
12	1.469	0.560	-0.241	0.506	-2.900	1.330	-0.770	1.084	0.202	0.088	-0.088	0.007	0.561
13	1.413	0.502	-0.097	0.493	-2.351	0.779	-0.371	0.685	0.201	0.015	0.011	0.010	0.480
14	1.245	0.652	-0.535	0.373	-1.539	0.657	-0.539	0.375	0.196	0.114	-0.114	0.010	0.443
15	0.649	0.251	-0.121	0.220	-2.612	0.976	-0.492	0.843	0.196	0.047	0.046	0.006	0.390
16	0.919	0.331	-0.261	0.203	-1.081	0.326	-0.151	0.289	0.181	0.032	0.031	0.010	0.427
17	1.620	0.602	-0.360	0.482	-3.120	1.234	-0.680	1.029	0.200	0.021	-0.020	0.005	0.488
18	1.616	0.446	-0.184	0.406	-1.771	0.478	-0.231	0.418	0.198	0.031	0.028	0.013	0.447
19	0.762	0.548	-0.508	0.205	-1.314	0.841	-0.734	0.410	0.190	0.120	-0.120	0.008	0.485
20	0.259	0.111	-0.002	0.111	-3.560	3.912	0.260	3.903	0.200	0.001	0.000	0.001	0.582
21	0.904	0.412	-0.376	0.169	-0.433	0.217	0.007	0.216	0.165	0.047	0.045	0.014	0.537
22	1.483	0.448	-0.207	0.397	-1.314	0.337	-0.084	0.327	0.184	0.065	0.064	0.013	0.485
23	0.793	0.621	-0.597	0.171	0.198	0.281	-0.062	0.274	0.175	0.067	-0.065	0.017	0.528
24	0.814	0.316	-0.156	0.275	-2.295	0.830	-0.425	0.713	0.195	0.036	0.035	0.007	0.429
25	1.819	0.521	-0.401	0.332	-2.010	0.499	-0.340	0.365	0.199	0.018	0.009	0.016	0.558
26	0.729	0.310	-0.262	0.166	-1.283	0.546	-0.403	0.369	0.189	0.032	-0.032	0.007	0.424
27	1.275	0.535	-0.385	0.371	-1.334	0.428	-0.254	0.344	0.183	0.025	0.023	0.010	0.572
28	0.634	0.298	-0.267	0.134	0.913	0.608	0.453	0.406	0.163	0.028	0.023	0.016	0.557
29	1.940	0.657	-0.620	0.216	-2.323	0.528	-0.363	0.383	0.192	0.053	0.052	0.011	0.615
30	1.377	0.416	-0.163	0.382	-1.098	0.275	-0.028	0.273	0.177	0.088	0.087	0.011	0.525
31	1.551	0.503	-0.379	0.331	-0.718	0.165	0.002	0.165	0.150	0.091	0.090	0.014	0.547

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.198	0.807	-0.773	0.235	-0.518	0.350	-0.288	0.198	0.182	0.090	-0.088	0.016	0.402
33	1.264	0.507	-0.416	0.290	-0.498	0.187	0.042	0.182	0.151	0.082	0.081	0.014	0.584
34	1.914	0.727	-0.696	0.209	-2.101	0.421	-0.351	0.232	0.193	0.026	0.023	0.012	0.602
35	1.921	0.579	-0.539	0.211	-2.148	0.436	-0.338	0.275	0.191	0.043	0.041	0.012	0.590
36	0.648	1.491	-1.482	0.165	3.243	2.172	1.803	1.211	0.139	0.031	-0.021	0.022	1.354
37	0.410	0.155	-0.100	0.119	-1.000	0.579	-0.220	0.535	0.196	0.016	-0.014	0.007	0.345
38	0.767	0.349	-0.253	0.240	-1.433	0.587	-0.373	0.453	0.192	0.009	0.002	0.009	0.506
39	0.574	0.404	-0.386	0.121	2.247	1.220	0.647	1.035	0.185	0.051	-0.046	0.022	0.526
40	0.866	0.516	-0.484	0.179	-0.642	0.462	-0.362	0.288	0.186	0.095	-0.094	0.010	0.397
41	0.646	0.285	-0.204	0.198	-2.710	1.249	-0.760	0.991	0.198	0.018	0.018	0.005	0.453
42	1.530	0.665	-0.510	0.426	-1.663	0.593	-0.423	0.415	0.195	0.038	-0.036	0.013	0.586
43	0.617	0.549	-0.533	0.132	0.280	0.483	-0.340	0.343	0.190	0.151	-0.150	0.014	0.511
44	0.445	0.224	-0.156	0.161	-3.090	1.807	-1.060	1.464	0.198	0.005	-0.003	0.004	0.435
45	1.885	0.948	-0.915	0.249	-1.957	0.611	-0.527	0.309	0.198	0.093	-0.092	0.016	0.583
46	1.121	0.609	-0.580	0.188	0.074	0.250	0.154	0.197	0.136	0.057	0.056	0.012	0.644
47	0.615	0.260	-0.185	0.182	3.089	1.643	1.249	1.068	0.162	0.049	0.042	0.025	0.584
48	0.643	0.288	-0.258	0.129	-0.530	0.316	0.040	0.314	0.179	0.069	0.069	0.010	0.525
49	0.678	0.169	-0.112	0.127	-0.174	0.308	0.156	0.266	0.180	0.051	0.050	0.012	0.357
50	0.352	0.117	0.042	0.110	1.438	0.795	-0.002	0.795	0.206	0.012	-0.005	0.011	0.304

Table E.9. Item-Level Evaluation Criteria Summary Statistics, Condition 2, Sample Size = 200

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.000	0.240	-0.240	0.000	-1.943	0.311	-0.193	0.244	0.200	0.070	0.070	0.000	0.409
2	1.000	0.090	-0.090	0.000	-0.568	0.337	-0.278	0.192	0.200	0.130	-0.130	0.000	0.307
3	1.000	0.370	-0.370	0.000	-2.301	0.500	-0.421	0.269	0.200	0.060	0.060	0.000	0.526
4	1.000	0.270	-0.270	0.000	-2.240	0.399	-0.290	0.274	0.200	0.050	0.050	0.000	0.424
5	1.000	0.210	0.210	0.000	-2.399	0.606	0.541	0.272	0.200	0.020	0.020	0.000	0.330
6	1.000	0.260	-0.260	0.000	-3.038	0.597	-0.508	0.314	0.200	0.020	0.020	0.000	0.377
7	1.000	0.670	-0.670	0.000	-1.840	0.524	-0.460	0.251	0.200	0.060	0.060	0.000	0.752
8	1.000	1.720	-1.720	0.000	-2.007	0.796	-0.757	0.245	0.200	0.070	0.070	0.000	1.233
9	1.000	2.650	-2.650	0.000	-2.883	1.471	-1.433	0.332	0.200	0.110	-0.110	0.000	1.163
10	1.000	0.640	-0.640	0.000	-3.003	0.844	-0.793	0.290	0.200	0.030	0.030	0.000	0.620
11	1.000	0.850	-0.850	0.000	-2.041	0.683	-0.631	0.262	0.200	0.030	0.030	0.000	0.822
12	1.000	0.710	-0.710	0.000	-3.241	1.169	-1.111	0.363	0.200	0.090	-0.090	0.000	0.635
13	1.000	0.510	-0.510	0.000	-2.618	0.701	-0.638	0.291	0.200	0.010	0.010	0.000	0.566
14	1.000	0.780	-0.780	0.000	-1.634	0.672	-0.634	0.224	0.200	0.110	-0.110	0.000	0.449
15	1.000	0.230	0.230	0.000	-1.633	0.544	0.487	0.242	0.200	0.050	0.050	0.000	0.326
16	1.000	0.180	-0.180	0.000	-0.886	0.212	0.044	0.208	0.200	0.050	0.050	0.000	0.313
17	1.000	0.980	-0.980	0.000	-3.761	1.388	-1.321	0.424	0.200	0.020	-0.020	0.000	0.703
18	1.000	0.800	-0.800	0.000	-2.205	0.723	-0.665	0.283	0.200	0.030	0.030	0.000	0.790
19	1.000	0.270	-0.270	0.000	-0.921	0.421	-0.341	0.246	0.200	0.110	-0.110	0.000	0.243
20	1.000	0.740	0.740	0.000	-1.036	2.847	2.784	0.598	0.200	0.000	0.000	0.000	1.195
21	1.000	0.280	-0.280	0.000	-0.256	0.272	0.185	0.200	0.200	0.080	0.080	0.000	0.452
22	1.000	0.690	-0.690	0.000	-1.539	0.411	-0.309	0.271	0.200	0.080	0.080	0.000	0.734
23	1.000	0.390	-0.390	0.000	0.298	0.232	0.038	0.229	0.200	0.040	-0.040	0.000	0.342
24	1.000	0.030	0.030	0.000	-1.742	0.286	0.128	0.255	0.200	0.040	0.040	0.000	0.198
25	1.000	1.220	-1.220	0.000	-2.703	1.071	-1.033	0.282	0.200	0.010	0.010	0.000	1.039
26	1.000	0.010	0.010	0.000	-0.865	0.202	0.015	0.201	0.200	0.020	-0.020	0.000	0.178
27	1.000	0.660	-0.660	0.000	-1.402	0.390	-0.322	0.220	0.200	0.040	0.040	0.000	0.648
28	1.000	0.100	0.100	0.000	0.857	0.460	0.397	0.232	0.200	0.060	0.060	0.000	0.290
29	1.000	1.560	-1.560	0.000	-3.244	1.347	-1.284	0.405	0.200	0.060	0.060	0.000	1.197
30	1.000	0.540	-0.540	0.000	-1.208	0.252	-0.138	0.211	0.200	0.110	0.110	0.000	0.700
31	1.000	0.930	-0.930	0.000	-0.806	0.236	-0.086	0.220	0.200	0.140	0.140	0.000	0.982

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.000	0.970	-0.970	0.000	-0.519	0.353	-0.289	0.202	0.200	0.070	-0.070	0.000	0.463
33	1.000	0.680	-0.680	0.000	-0.443	0.232	0.097	0.211	0.200	0.130	0.130	0.000	0.814
34	1.000	1.610	-1.610	0.000	-2.929	1.217	-1.179	0.304	0.200	0.030	0.030	0.000	1.217
35	1.000	1.460	-1.460	0.000	-2.991	1.228	-1.181	0.336	0.200	0.050	0.050	0.000	1.180
36	1.000	1.130	-1.130	0.000	2.632	1.266	1.192	0.425	0.200	0.040	0.040	0.000	1.158
37	1.000	0.490	0.490	0.000	-0.259	0.560	0.521	0.205	0.200	0.010	-0.010	0.000	0.726
38	1.000	0.020	-0.020	0.000	-1.013	0.241	0.047	0.237	0.200	0.010	0.010	0.000	0.225
39	1.000	0.040	0.040	0.000	1.641	0.787	0.041	0.785	0.200	0.030	-0.030	0.000	0.342
40	1.000	0.350	-0.350	0.000	-0.470	0.293	-0.190	0.223	0.200	0.080	-0.080	0.000	0.264
41	1.000	0.150	0.150	0.000	-1.684	0.360	0.266	0.242	0.200	0.020	0.020	0.000	0.259
42	1.000	1.040	-1.040	0.000	-1.983	0.799	-0.743	0.295	0.200	0.030	-0.030	0.000	0.819
43	1.000	0.150	-0.150	0.000	0.337	0.368	-0.283	0.235	0.200	0.140	-0.140	0.000	0.468
44	1.000	0.400	0.400	0.000	-1.250	0.816	0.780	0.241	0.200	0.000	0.000	0.000	0.610
45	1.000	1.800	-1.800	0.000	-2.709	1.325	-1.279	0.347	0.200	0.090	-0.090	0.000	1.069
46	1.000	0.700	-0.700	0.000	0.246	0.397	0.326	0.227	0.200	0.120	0.120	0.000	0.825
47	1.000	0.200	0.200	0.000	2.386	0.659	0.546	0.368	0.200	0.080	0.080	0.000	0.412
48	1.000	0.100	0.100	0.000	-0.204	0.424	0.366	0.215	0.200	0.090	0.090	0.000	0.224
49	1.000	0.210	0.210	0.000	0.027	0.407	0.357	0.196	0.200	0.070	0.070	0.000	0.228
50	1.000	0.690	0.690	0.000	0.765	0.706	-0.675	0.206	0.200	0.010	-0.010	0.000	0.988

Table E.10. Item-Level Evaluation Criteria Summary Statistics, Condition 3, Sample Size = 200

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.108	0.431	-0.132	0.410	-2.088	0.855	-0.338	0.785	0.200	0.070	0.070	0.000	0.509
2	0.662	0.474	-0.428	0.203	-1.006	0.832	-0.716	0.424	0.200	0.130	-0.130	0.000	0.429
3	1.245	0.494	-0.125	0.478	-2.271	0.795	-0.391	0.692	0.200	0.060	0.060	0.000	0.543
4	1.073	0.446	-0.197	0.400	-2.438	0.985	-0.488	0.856	0.200	0.050	0.050	0.000	0.529
5	0.678	0.346	-0.112	0.328	-3.991	2.033	-1.051	1.740	0.200	0.020	0.020	0.000	0.380
6	1.058	0.547	-0.202	0.509	-3.226	2.765	-0.696	2.676	0.200	0.020	0.020	0.000	0.907
7	1.526	0.446	-0.144	0.422	-1.531	0.369	-0.151	0.336	0.200	0.060	0.060	0.000	0.438
8	1.935	0.809	-0.785	0.194	-1.444	0.286	-0.194	0.210	0.200	0.070	0.070	0.000	0.609
9	1.952	1.707	-1.698	0.173	-2.058	0.677	-0.608	0.297	0.200	0.110	-0.110	0.000	0.636
10	1.497	0.498	-0.143	0.477	-2.607	0.849	-0.397	0.750	0.200	0.030	0.030	0.000	0.447
11	1.682	0.427	-0.168	0.393	-1.601	0.399	-0.191	0.351	0.200	0.030	0.030	0.000	0.489
12	1.469	0.560	-0.241	0.506	-2.910	1.333	-0.780	1.081	0.200	0.090	-0.090	0.000	0.563
13	1.413	0.502	-0.097	0.493	-2.359	0.780	-0.379	0.682	0.200	0.010	0.010	0.000	0.480
14	1.245	0.652	-0.535	0.373	-1.539	0.655	-0.539	0.373	0.200	0.110	-0.110	0.000	0.446
15	0.649	0.251	-0.121	0.220	-2.606	0.980	-0.486	0.851	0.200	0.050	0.050	0.000	0.389
16	0.919	0.331	-0.261	0.203	-1.043	0.321	-0.113	0.300	0.200	0.050	0.050	0.000	0.446
17	1.620	0.602	-0.360	0.482	-3.127	1.236	-0.687	1.027	0.200	0.020	-0.020	0.000	0.490
18	1.616	0.446	-0.184	0.406	-1.775	0.478	-0.235	0.416	0.200	0.030	0.030	0.000	0.449
19	0.762	0.548	-0.508	0.205	-1.293	0.825	-0.713	0.416	0.200	0.110	-0.110	0.000	0.487
20	0.259	0.111	-0.002	0.111	-3.372	4.118	0.448	4.094	0.200	0.000	0.000	0.000	0.644
21	0.904	0.412	-0.376	0.169	-0.351	0.249	0.089	0.232	0.200	0.080	0.080	0.000	0.590
22	1.483	0.448	-0.207	0.397	-1.297	0.338	-0.067	0.331	0.200	0.080	0.080	0.000	0.500
23	0.793	0.621	-0.597	0.171	0.274	0.275	0.014	0.275	0.200	0.040	-0.040	0.000	0.541
24	0.814	0.316	-0.156	0.275	-2.289	0.830	-0.419	0.717	0.200	0.040	0.040	0.000	0.430
25	1.819	0.521	-0.401	0.332	-2.018	0.502	-0.348	0.362	0.200	0.010	0.010	0.000	0.563
26	0.729	0.310	-0.262	0.166	-1.258	0.533	-0.378	0.376	0.200	0.020	-0.020	0.000	0.431
27	1.275	0.535	-0.385	0.371	-1.312	0.419	-0.232	0.349	0.200	0.040	0.040	0.000	0.583
28	0.634	0.298	-0.267	0.134	1.092	0.753	0.632	0.410	0.200	0.060	0.060	0.000	0.609
29	1.940	0.657	-0.620	0.216	-2.324	0.528	-0.364	0.382	0.200	0.060	0.060	0.000	0.626
30	1.377	0.416	-0.163	0.382	-1.068	0.277	0.002	0.277	0.200	0.110	0.110	0.000	0.553
31	1.551	0.503	-0.379	0.331	-0.654	0.186	0.066	0.174	0.200	0.140	0.140	0.000	0.656

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.198	0.807	-0.773	0.235	-0.486	0.328	-0.256	0.206	0.200	0.070	-0.070	0.000	0.402
33	1.264	0.507	-0.416	0.290	-0.416	0.230	0.124	0.194	0.200	0.130	0.130	0.000	0.679
34	1.914	0.727	-0.696	0.209	-2.103	0.423	-0.353	0.234	0.200	0.030	0.030	0.000	0.615
35	1.921	0.579	-0.539	0.211	-2.148	0.438	-0.338	0.278	0.200	0.050	0.050	0.000	0.600
36	0.648	1.491	-1.482	0.165	3.964	2.930	2.524	1.488	0.200	0.040	0.040	0.000	1.497
37	0.410	0.155	-0.100	0.119	-0.986	0.584	-0.206	0.546	0.200	0.010	-0.010	0.000	0.341
38	0.767	0.349	-0.253	0.240	-1.414	0.582	-0.354	0.461	0.200	0.010	0.010	0.000	0.509
39	0.574	0.404	-0.386	0.121	2.341	1.183	0.741	0.922	0.200	0.030	-0.030	0.000	0.534
40	0.866	0.516	-0.484	0.179	-0.611	0.445	-0.331	0.298	0.200	0.080	-0.080	0.000	0.395
41	0.646	0.285	-0.204	0.198	-2.706	1.250	-0.756	0.996	0.200	0.020	0.020	0.000	0.453
42	1.530	0.665	-0.510	0.426	-1.661	0.591	-0.421	0.415	0.200	0.030	-0.030	0.000	0.590
43	0.617	0.549	-0.533	0.132	0.317	0.457	-0.303	0.341	0.200	0.140	-0.140	0.000	0.502
44	0.445	0.224	-0.156	0.161	-3.086	1.813	-1.056	1.474	0.200	0.000	0.000	0.000	0.434
45	1.885	0.948	-0.915	0.249	-1.964	0.617	-0.534	0.308	0.200	0.090	-0.090	0.000	0.589
46	1.121	0.609	-0.580	0.188	0.217	0.369	0.297	0.219	0.200	0.120	0.120	0.000	0.784
47	0.615	0.260	-0.185	0.182	3.481	1.878	1.641	0.913	0.200	0.080	0.080	0.000	0.679
48	0.643	0.288	-0.258	0.129	-0.460	0.348	0.110	0.330	0.200	0.090	0.090	0.000	0.551
49	0.678	0.169	-0.112	0.127	-0.107	0.358	0.223	0.280	0.200	0.070	0.070	0.000	0.374
50	0.352	0.117	0.042	0.110	1.353	0.706	-0.088	0.701	0.200	0.010	-0.010	0.000	0.298

Table E.11. Item-Level Evaluation Criteria Summary Statistics, Condition 4, Sample Size = 200

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.108	0.431	-0.132	0.410	-1.478	0.510	0.272	0.431	0.200	0.070	0.070	0.000	0.463
2	0.662	0.473	-0.428	0.202	-0.662	0.447	-0.372	0.247	0.200	0.130	-0.130	0.000	0.432
3	1.245	0.494	-0.125	0.478	-1.598	0.478	0.282	0.385	0.200	0.060	0.060	0.000	0.520
4	1.073	0.446	-0.197	0.400	-1.690	0.540	0.260	0.473	0.200	0.050	0.050	0.000	0.469
5	0.677	0.347	-0.113	0.328	-2.555	1.045	0.385	0.972	0.200	0.020	0.020	0.000	0.637
6	1.058	0.547	-0.202	0.509	-2.415	1.059	0.115	1.052	0.200	0.020	0.020	0.000	0.649
7	1.526	0.446	-0.144	0.422	-1.145	0.312	0.235	0.204	0.200	0.060	0.060	0.000	0.380
8	1.935	0.809	-0.785	0.193	-1.059	0.242	0.191	0.149	0.200	0.070	0.070	0.000	0.451
9	1.952	1.707	-1.698	0.172	-1.429	0.219	0.021	0.218	0.200	0.110	-0.110	0.000	0.441
10	1.497	0.498	-0.143	0.477	-1.813	0.586	0.397	0.431	0.200	0.030	0.030	0.000	0.638
11	1.682	0.428	-0.168	0.393	-1.180	0.322	0.230	0.226	0.200	0.030	0.030	0.000	0.436
12	1.469	0.560	-0.241	0.506	-1.997	0.615	0.134	0.600	0.200	0.090	-0.090	0.000	0.641
13	1.413	0.502	-0.097	0.493	-1.652	0.510	0.328	0.391	0.200	0.010	0.010	0.000	0.590
14	1.246	0.652	-0.534	0.373	-1.128	0.247	-0.128	0.212	0.200	0.110	-0.110	0.000	0.390
15	0.649	0.251	-0.121	0.220	-1.730	0.620	0.390	0.482	0.200	0.050	0.050	0.000	0.463
16	0.919	0.331	-0.261	0.203	-0.782	0.230	0.148	0.176	0.200	0.050	0.050	0.000	0.390
17	1.620	0.601	-0.360	0.482	-2.130	0.658	0.310	0.580	0.200	0.020	-0.020	0.000	0.709
18	1.616	0.446	-0.184	0.406	-1.298	0.351	0.242	0.254	0.200	0.030	0.030	0.000	0.434
19	0.762	0.548	-0.508	0.205	-0.893	0.401	-0.313	0.251	0.200	0.110	-0.110	0.000	0.434
20	0.259	0.110	-0.001	0.110	-2.252	1.992	1.568	1.229	0.200	0.000	0.000	0.000	0.635
21	0.904	0.412	-0.376	0.169	-0.321	0.189	0.119	0.147	0.200	0.080	0.080	0.000	0.566
22	1.483	0.448	-0.207	0.397	-0.994	0.311	0.236	0.202	0.200	0.080	0.080	0.000	0.443
23	0.794	0.620	-0.596	0.171	0.151	0.237	-0.109	0.210	0.200	0.040	-0.040	0.000	0.503
24	0.815	0.315	-0.155	0.275	-1.558	0.510	0.312	0.404	0.200	0.040	0.040	0.000	0.437
25	1.819	0.521	-0.401	0.332	-1.435	0.337	0.235	0.242	0.200	0.010	0.010	0.000	0.461
26	0.729	0.310	-0.262	0.166	-0.867	0.222	0.013	0.222	0.200	0.020	-0.020	0.000	0.390
27	1.275	0.535	-0.386	0.371	-0.985	0.223	0.095	0.202	0.200	0.040	0.040	0.000	0.468
28	0.633	0.299	-0.267	0.133	0.627	0.337	0.167	0.293	0.200	0.060	0.060	0.000	0.569
29	1.940	0.657	-0.620	0.216	-1.601	0.443	0.359	0.260	0.200	0.060	0.060	0.000	0.550
30	1.377	0.416	-0.163	0.382	-0.846	0.279	0.225	0.165	0.200	0.110	0.110	0.000	0.470
31	1.551	0.503	-0.379	0.331	-0.581	0.179	0.139	0.112	0.200	0.140	0.140	0.000	0.615

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.198	0.807	-0.772	0.235	-0.405	0.226	-0.175	0.143	0.200	0.070	-0.070	0.000	0.375
33	1.264	0.507	-0.416	0.289	-0.412	0.182	0.128	0.129	0.200	0.130	0.130	0.000	0.663
34	1.914	0.727	-0.696	0.209	-1.474	0.323	0.276	0.167	0.200	0.030	0.030	0.000	0.449
35	1.921	0.579	-0.539	0.210	-1.507	0.363	0.303	0.200	0.200	0.050	0.050	0.000	0.485
36	0.648	1.491	-1.482	0.166	2.195	1.065	0.755	0.750	0.200	0.040	0.040	0.000	1.096
37	0.409	0.155	-0.101	0.118	-0.601	0.384	0.180	0.339	0.200	0.010	-0.010	0.000	0.351
38	0.768	0.348	-0.252	0.240	-0.981	0.280	0.079	0.269	0.200	0.010	0.010	0.000	0.441
39	0.575	0.404	-0.386	0.121	1.614	0.938	0.014	0.938	0.200	0.030	-0.030	0.000	0.478
40	0.866	0.516	-0.484	0.179	-0.464	0.266	-0.184	0.192	0.200	0.080	-0.080	0.000	0.378
41	0.645	0.285	-0.205	0.198	-1.783	0.596	0.167	0.572	0.200	0.020	0.020	0.000	0.456
42	1.529	0.665	-0.511	0.426	-1.219	0.253	0.021	0.252	0.200	0.030	-0.030	0.000	0.419
43	0.618	0.549	-0.533	0.132	0.235	0.466	-0.385	0.263	0.200	0.140	-0.140	0.000	0.444
44	0.444	0.224	-0.156	0.161	-1.966	0.867	0.064	0.864	0.200	0.000	0.000	0.000	0.508
45	1.885	0.948	-0.915	0.249	-1.391	0.229	0.039	0.226	0.200	0.090	-0.090	0.000	0.404
46	1.121	0.609	-0.579	0.188	0.011	0.175	0.091	0.149	0.200	0.120	0.120	0.000	0.791
47	0.615	0.260	-0.185	0.182	2.085	0.806	0.245	0.768	0.200	0.080	0.080	0.000	0.661
48	0.643	0.288	-0.257	0.129	-0.363	0.295	0.208	0.210	0.200	0.090	0.090	0.000	0.521
49	0.677	0.170	-0.113	0.126	-0.138	0.261	0.193	0.177	0.200	0.070	0.070	0.000	0.348
50	0.351	0.117	0.041	0.109	1.198	0.813	-0.242	0.776	0.200	0.010	-0.010	0.000	0.314

Table E.12. Item-Level Evaluation Criteria Summary Statistics, Condition 5, Sample Size = 200

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.087	0.318	-0.153	0.279	-1.906	0.597	-0.156	0.576	0.193	0.063	0.063	0.008	0.408
2	0.828	0.388	-0.262	0.286	-0.752	0.596	-0.462	0.376	0.205	0.126	-0.125	0.016	0.398
3	1.170	0.367	-0.200	0.308	-2.114	0.545	-0.234	0.492	0.195	0.056	0.055	0.008	0.447
4	1.092	0.360	-0.178	0.313	-2.211	0.755	-0.261	0.708	0.196	0.046	0.046	0.008	0.434
5	0.711	0.263	-0.079	0.251	-3.454	1.342	-0.514	1.240	0.199	0.019	0.019	0.004	0.332
6	1.019	0.424	-0.241	0.349	-3.249	1.393	-0.719	1.193	0.199	0.020	0.019	0.005	0.491
7	1.506	0.430	-0.164	0.397	-1.419	0.295	-0.039	0.292	0.192	0.054	0.052	0.014	0.373
8	2.365	0.727	-0.356	0.635	-1.254	0.243	-0.004	0.243	0.184	0.058	0.054	0.022	0.456
9	2.361	1.400	-1.289	0.547	-1.820	0.531	-0.370	0.381	0.200	0.112	-0.110	0.022	0.493
10	1.330	0.476	-0.310	0.361	-2.541	0.657	-0.331	0.568	0.198	0.028	0.028	0.007	0.424
11	1.669	0.483	-0.181	0.448	-1.491	0.328	-0.081	0.318	0.194	0.029	0.024	0.016	0.454
12	1.320	0.540	-0.390	0.374	-2.793	1.078	-0.663	0.850	0.202	0.089	-0.088	0.005	0.524
13	1.281	0.386	-0.229	0.311	-2.241	0.536	-0.261	0.469	0.198	0.011	0.008	0.007	0.406
14	1.343	0.534	-0.437	0.306	-1.324	0.440	-0.324	0.297	0.202	0.109	-0.108	0.015	0.351
15	0.677	0.234	-0.093	0.215	-2.424	0.876	-0.304	0.821	0.197	0.047	0.047	0.006	0.359
16	1.103	0.259	-0.077	0.247	-0.847	0.267	0.083	0.254	0.189	0.042	0.039	0.015	0.316
17	1.402	0.694	-0.578	0.384	-3.104	1.079	-0.664	0.851	0.200	0.020	-0.020	0.004	0.499
18	1.580	0.410	-0.220	0.346	-1.641	0.338	-0.101	0.323	0.197	0.031	0.027	0.015	0.410
19	0.949	0.403	-0.321	0.244	-0.996	0.556	-0.416	0.370	0.203	0.108	-0.107	0.014	0.380
20	0.291	0.117	0.031	0.113	-3.769	1.662	0.052	1.661	0.200	0.002	0.000	0.002	0.283
21	1.267	0.273	-0.013	0.273	-0.221	0.284	0.219	0.181	0.180	0.065	0.060	0.023	0.343
22	1.517	0.377	-0.173	0.335	-1.179	0.311	0.051	0.307	0.186	0.068	0.066	0.017	0.428
23	1.152	0.470	-0.238	0.405	0.312	0.236	0.052	0.231	0.198	0.050	-0.042	0.027	0.406
24	0.858	0.263	-0.112	0.237	-2.063	0.638	-0.193	0.608	0.196	0.037	0.036	0.007	0.363
25	1.811	0.632	-0.409	0.482	-1.905	0.420	-0.235	0.348	0.198	0.014	0.008	0.012	0.521
26	0.898	0.221	-0.092	0.201	-0.973	0.311	-0.093	0.297	0.201	0.024	-0.020	0.013	0.302
27	1.508	0.389	-0.153	0.357	-1.068	0.265	0.012	0.265	0.193	0.037	0.033	0.017	0.402
28	0.906	0.274	0.006	0.274	0.844	0.515	0.384	0.343	0.175	0.040	0.035	0.019	0.430
29	1.941	0.756	-0.620	0.434	-2.231	0.568	-0.271	0.499	0.194	0.054	0.054	0.010	0.544
30	1.464	0.329	-0.076	0.321	-0.943	0.256	0.127	0.222	0.179	0.091	0.089	0.016	0.428
31	1.875	0.434	-0.056	0.431	-0.556	0.230	0.164	0.161	0.156	0.097	0.096	0.018	0.442

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.618	0.530	-0.352	0.396	-0.296	0.216	-0.066	0.205	0.207	0.071	-0.063	0.034	0.320
33	1.613	0.335	-0.067	0.328	-0.335	0.265	0.205	0.168	0.159	0.090	0.089	0.018	0.424
34	1.957	0.775	-0.654	0.417	-1.975	0.364	-0.225	0.286	0.195	0.027	0.025	0.011	0.530
35	1.933	0.675	-0.527	0.422	-2.040	0.413	-0.230	0.343	0.192	0.044	0.042	0.011	0.562
36	1.498	0.778	-0.632	0.455	1.957	0.800	0.517	0.610	0.152	0.026	-0.008	0.025	0.738
37	0.472	0.159	-0.039	0.154	-0.806	0.488	-0.026	0.487	0.199	0.014	-0.011	0.008	0.340
38	0.911	0.291	-0.109	0.270	-1.149	0.409	-0.089	0.399	0.201	0.022	0.011	0.019	0.406
39	0.721	0.396	-0.239	0.316	2.087	1.032	0.487	0.910	0.190	0.048	-0.041	0.025	0.517
40	1.061	0.376	-0.290	0.240	-0.446	0.318	-0.166	0.271	0.200	0.082	-0.080	0.019	0.354
41	0.725	0.241	-0.125	0.207	-2.337	0.913	-0.387	0.827	0.199	0.020	0.019	0.005	0.376
42	1.642	0.588	-0.398	0.433	-1.453	0.395	-0.213	0.333	0.199	0.036	-0.031	0.018	0.465
43	0.779	0.448	-0.371	0.250	0.364	0.398	-0.256	0.305	0.203	0.138	-0.137	0.021	0.497
44	0.516	0.208	-0.084	0.191	-2.584	1.292	-0.554	1.167	0.199	0.004	-0.001	0.004	0.407
45	2.016	0.925	-0.784	0.490	-1.808	0.529	-0.378	0.370	0.202	0.090	-0.088	0.015	0.510
46	1.743	0.351	0.043	0.348	0.170	0.297	0.250	0.160	0.149	0.072	0.069	0.018	0.412
47	0.784	0.281	-0.016	0.281	2.765	1.605	0.925	1.311	0.163	0.047	0.043	0.020	0.524
48	0.830	0.216	-0.070	0.205	-0.335	0.371	0.235	0.288	0.188	0.079	0.078	0.011	0.385
49	0.802	0.192	0.012	0.192	-0.065	0.353	0.265	0.234	0.186	0.059	0.056	0.016	0.310
50	0.347	0.167	0.037	0.163	1.532	0.866	0.092	0.861	0.204	0.013	-0.007	0.011	0.341

Table E.13. Item-Level Evaluation Criteria Summary Statistics, Condition 0, Sample Size = 150

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.270	0.397	0.030	0.396	-1.777	0.657	-0.027	0.657	0.195	0.065	0.065	0.007	0.420
2	0.936	0.419	-0.154	0.390	-0.746	0.665	-0.456	0.484	0.206	0.125	-0.124	0.015	0.466
3	1.434	0.464	0.064	0.459	-1.903	0.473	-0.023	0.473	0.197	0.058	0.057	0.008	0.431
4	1.339	0.616	0.069	0.612	-2.051	0.775	-0.101	0.769	0.198	0.049	0.048	0.010	0.525
5	0.832	0.342	0.042	0.339	-3.072	1.143	-0.132	1.136	0.199	0.019	0.019	0.004	0.428
6	1.268	0.570	0.008	0.570	-2.790	1.835	-0.260	1.816	0.199	0.020	0.019	0.005	0.650
7	1.857	0.675	0.187	0.648	-1.309	0.313	0.071	0.305	0.197	0.060	0.057	0.017	0.386
8	2.998	1.081	0.278	1.044	-1.144	0.275	0.106	0.253	0.192	0.067	0.062	0.024	0.475
9	3.110	0.948	-0.540	0.779	-1.618	0.363	-0.168	0.321	0.205	0.107	-0.105	0.021	0.469
10	1.622	0.539	-0.018	0.539	-2.322	0.677	-0.112	0.667	0.199	0.030	0.029	0.006	0.453
11	2.095	0.699	0.245	0.654	-1.332	0.294	0.078	0.284	0.198	0.032	0.028	0.017	0.418
12	1.714	0.615	0.004	0.615	-2.447	0.951	-0.317	0.896	0.202	0.088	-0.088	0.008	0.524
13	1.593	0.561	0.083	0.555	-2.096	0.739	-0.116	0.729	0.199	0.012	0.009	0.008	0.485
14	1.599	0.575	-0.181	0.545	-1.263	0.443	-0.263	0.356	0.203	0.109	-0.107	0.021	0.427
15	0.759	0.275	-0.011	0.274	-2.298	0.951	-0.178	0.935	0.198	0.048	0.048	0.006	0.383
16	1.242	0.284	0.062	0.277	-0.808	0.262	0.122	0.232	0.191	0.043	0.041	0.013	0.297
17	1.925	0.647	-0.056	0.645	-2.525	1.683	-0.085	1.681	0.202	0.019	-0.018	0.006	0.524
18	1.983	0.599	0.183	0.571	-1.469	0.328	0.071	0.320	0.200	0.035	0.030	0.018	0.403
19	1.094	0.423	-0.176	0.384	-0.957	0.576	-0.377	0.436	0.204	0.108	-0.107	0.016	0.442
20	0.345	0.168	0.085	0.145	-3.045	2.199	0.775	2.058	0.200	0.002	0.000	0.002	0.425
21	1.439	0.412	0.159	0.380	-0.232	0.271	0.208	0.173	0.186	0.069	0.066	0.020	0.351
22	1.782	0.514	0.092	0.506	-1.092	0.333	0.138	0.303	0.191	0.073	0.071	0.017	0.421
23	1.308	0.496	-0.083	0.490	0.271	0.243	0.011	0.243	0.201	0.047	-0.039	0.027	0.389
24	1.019	0.377	0.049	0.374	-1.892	0.687	-0.022	0.686	0.198	0.039	0.038	0.010	0.452
25	2.310	0.792	0.090	0.787	-1.726	0.410	-0.056	0.406	0.203	0.019	0.013	0.014	0.500
26	1.063	0.366	0.073	0.358	-0.923	0.380	-0.043	0.378	0.204	0.025	-0.016	0.019	0.372
27	1.830	0.624	0.170	0.600	-0.971	0.298	0.109	0.277	0.198	0.044	0.038	0.021	0.414
28	0.999	0.392	0.099	0.379	0.785	0.512	0.325	0.395	0.184	0.048	0.044	0.019	0.465
29	2.655	0.686	0.095	0.679	-1.955	0.435	0.005	0.435	0.196	0.057	0.056	0.011	0.490
30	1.704	0.548	0.164	0.523	-0.885	0.304	0.185	0.241	0.187	0.098	0.097	0.017	0.411
31	2.224	0.752	0.294	0.692	-0.531	0.249	0.189	0.162	0.166	0.108	0.106	0.018	0.461

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.834	0.623	-0.136	0.608	-0.299	0.201	-0.069	0.188	0.208	0.068	-0.062	0.029	0.314
33	1.803	0.459	0.123	0.442	-0.327	0.279	0.213	0.179	0.168	0.099	0.098	0.017	0.424
34	2.640	0.690	0.030	0.689	-1.740	0.264	0.010	0.263	0.199	0.032	0.029	0.012	0.410
35	2.564	0.742	0.104	0.735	-1.822	0.381	-0.012	0.381	0.195	0.047	0.045	0.014	0.487
36	1.712	0.686	-0.419	0.543	1.808	0.650	0.368	0.536	0.160	0.025	0.000	0.025	0.664
37	0.537	0.189	0.027	0.187	-0.792	0.517	-0.012	0.517	0.200	0.013	-0.010	0.008	0.358
38	1.065	0.349	0.045	0.346	-1.065	0.397	-0.005	0.397	0.202	0.022	0.012	0.019	0.409
39	0.856	0.398	-0.104	0.385	1.884	0.934	0.284	0.890	0.198	0.039	-0.032	0.023	0.506
40	1.188	0.353	-0.162	0.313	-0.441	0.311	-0.161	0.267	0.201	0.081	-0.079	0.017	0.381
41	0.849	0.269	-0.001	0.269	-2.080	0.772	-0.130	0.761	0.199	0.020	0.019	0.005	0.357
42	2.026	0.657	-0.015	0.657	-1.337	0.415	-0.097	0.404	0.203	0.035	-0.028	0.021	0.499
43	0.898	0.425	-0.252	0.343	0.329	0.452	-0.291	0.346	0.208	0.134	-0.132	0.025	0.507
44	0.590	0.244	-0.010	0.244	-2.377	1.160	-0.347	1.107	0.200	0.005	0.000	0.005	0.414
45	2.641	0.782	-0.159	0.766	-1.625	0.476	-0.195	0.434	0.207	0.085	-0.083	0.016	0.457
46	1.951	0.535	0.251	0.473	0.137	0.280	0.217	0.178	0.161	0.083	0.081	0.021	0.418
47	0.941	0.410	0.141	0.385	2.550	1.416	0.710	1.226	0.174	0.059	0.054	0.023	0.585
48	0.909	0.289	0.009	0.289	-0.327	0.391	0.243	0.306	0.192	0.083	0.082	0.011	0.426
49	0.918	0.299	0.128	0.270	-0.063	0.366	0.268	0.249	0.191	0.063	0.061	0.014	0.351
50	0.418	0.211	0.108	0.182	1.387	0.969	-0.053	0.967	0.206	0.011	-0.004	0.010	0.408

Table E.14. Item-Level Evaluation Criteria Summary Statistics, Condition 1, Sample Size = 150

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.136	0.491	-0.104	0.480	-2.154	1.141	-0.404	1.067	0.196	0.066	0.066	0.007	0.549
2	0.655	0.496	-0.435	0.239	-1.087	1.043	-0.797	0.673	0.193	0.137	-0.137	0.010	0.501
3	1.276	0.503	-0.094	0.494	-2.275	0.893	-0.395	0.801	0.199	0.059	0.059	0.009	0.539
4	1.098	0.520	-0.172	0.491	-2.544	1.275	-0.594	1.128	0.196	0.047	0.046	0.007	0.607
5	0.677	0.359	-0.113	0.341	-3.564	2.824	-0.624	2.754	0.199	0.019	0.019	0.003	0.590
6	1.079	0.598	-0.181	0.570	-2.472	4.972	0.058	4.972	0.199	0.020	0.019	0.005	1.358
7	1.534	0.472	-0.137	0.452	-1.560	0.426	-0.180	0.386	0.194	0.056	0.054	0.013	0.493
8	1.865	0.905	-0.855	0.294	-1.483	0.386	-0.233	0.307	0.182	0.054	0.052	0.017	0.651
9	1.957	1.700	-1.693	0.151	-2.047	0.661	-0.597	0.283	0.194	0.117	-0.117	0.012	0.644
10	1.456	0.562	-0.185	0.531	-2.760	1.224	-0.550	1.094	0.199	0.030	0.029	0.006	0.523
11	1.696	0.427	-0.154	0.398	-1.596	0.428	-0.186	0.386	0.195	0.028	0.025	0.014	0.475
12	1.452	0.579	-0.258	0.519	-2.684	2.185	-0.554	2.113	0.202	0.089	-0.088	0.006	0.739
13	1.413	0.561	-0.097	0.553	-2.523	1.245	-0.543	1.121	0.201	0.014	0.011	0.008	0.564
14	1.285	0.668	-0.495	0.448	-1.590	0.790	-0.590	0.525	0.198	0.113	-0.113	0.012	0.524
15	0.641	0.281	-0.129	0.249	-2.757	1.328	-0.637	1.166	0.197	0.047	0.047	0.005	0.442
16	0.941	0.333	-0.239	0.231	-1.063	0.333	-0.133	0.306	0.184	0.035	0.034	0.008	0.432
17	1.602	0.665	-0.378	0.548	-2.406	5.538	0.034	5.537	0.201	0.020	-0.019	0.004	0.617
18	1.624	0.452	-0.176	0.417	-1.778	0.548	-0.238	0.493	0.196	0.030	0.026	0.014	0.467
19	0.786	0.560	-0.484	0.282	-1.341	0.982	-0.761	0.619	0.192	0.118	-0.118	0.008	0.541
20	0.262	0.123	0.002	0.123	-3.477	4.015	0.343	4.000	0.200	0.001	0.000	0.001	0.656
21	0.922	0.408	-0.358	0.196	-0.406	0.229	0.034	0.227	0.170	0.052	0.050	0.012	0.557
22	1.467	0.483	-0.223	0.428	-1.325	0.402	-0.095	0.391	0.188	0.069	0.068	0.011	0.534
23	0.793	0.627	-0.598	0.189	0.237	0.320	-0.023	0.319	0.178	0.064	-0.062	0.016	0.559
24	0.848	0.363	-0.122	0.342	-2.327	1.108	-0.457	1.010	0.196	0.037	0.036	0.006	0.496
25	1.760	0.612	-0.460	0.403	-2.095	0.730	-0.425	0.593	0.199	0.015	0.009	0.012	0.647
26	0.760	0.329	-0.230	0.234	-1.295	0.659	-0.415	0.513	0.191	0.030	-0.029	0.007	0.490
27	1.302	0.576	-0.358	0.451	-1.320	0.469	-0.240	0.402	0.187	0.029	0.027	0.012	0.600
28	0.633	0.307	-0.267	0.151	0.955	0.691	0.495	0.483	0.170	0.033	0.030	0.015	0.593
29	1.916	0.690	-0.644	0.248	-2.359	0.674	-0.399	0.543	0.193	0.054	0.053	0.009	0.672
30	1.369	0.442	-0.171	0.408	-1.097	0.324	-0.027	0.323	0.182	0.093	0.092	0.011	0.567
31	1.566	0.539	-0.364	0.397	-0.701	0.195	0.019	0.194	0.160	0.101	0.100	0.014	0.608

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.188	0.832	-0.782	0.283	-0.498	0.339	-0.268	0.208	0.183	0.088	-0.087	0.014	0.429
33	1.264	0.524	-0.416	0.318	-0.465	0.224	0.075	0.211	0.160	0.091	0.090	0.014	0.612
34	1.903	0.751	-0.707	0.254	-2.120	0.467	-0.370	0.285	0.194	0.026	0.024	0.011	0.645
35	1.897	0.637	-0.563	0.297	-2.203	0.659	-0.393	0.529	0.193	0.044	0.043	0.011	0.638
36	0.684	1.474	-1.446	0.287	3.417	2.434	1.977	1.419	0.146	0.026	-0.014	0.021	1.381
37	0.414	0.174	-0.096	0.145	-1.106	0.867	-0.326	0.803	0.196	0.015	-0.014	0.007	0.392
38	0.801	0.336	-0.219	0.255	-1.406	0.611	-0.346	0.504	0.193	0.009	0.003	0.008	0.516
39	0.567	0.425	-0.393	0.160	2.404	1.689	0.804	1.485	0.184	0.050	-0.046	0.020	0.585
40	0.868	0.523	-0.482	0.203	-0.635	0.489	-0.355	0.337	0.188	0.092	-0.092	0.010	0.444
41	0.668	0.294	-0.182	0.230	-2.661	1.286	-0.711	1.072	0.198	0.019	0.018	0.004	0.454
42	1.539	0.680	-0.501	0.461	-1.705	0.802	-0.465	0.653	0.194	0.038	-0.036	0.013	0.652
43	0.617	0.557	-0.533	0.161	0.308	0.526	-0.312	0.423	0.191	0.150	-0.149	0.012	0.543
44	0.446	0.236	-0.154	0.179	-2.780	2.599	-0.750	2.489	0.198	0.005	-0.002	0.004	0.659
45	1.854	1.002	-0.946	0.329	-2.039	0.882	-0.609	0.638	0.197	0.093	-0.093	0.013	0.629
46	1.139	0.613	-0.561	0.249	0.111	0.303	0.191	0.236	0.147	0.068	0.067	0.014	0.692
47	0.595	0.268	-0.205	0.174	3.745	3.774	1.905	3.258	0.165	0.051	0.045	0.024	0.666
48	0.638	0.314	-0.262	0.172	-0.500	0.379	0.070	0.373	0.183	0.073	0.073	0.009	0.589
49	0.696	0.184	-0.094	0.158	-0.134	0.366	0.196	0.309	0.182	0.054	0.052	0.011	0.393
50	0.347	0.132	0.037	0.127	1.612	1.249	0.172	1.237	0.202	0.012	-0.008	0.009	0.341

Table E.15. Item-Level Evaluation Criteria Summary Statistics, Condition 2, Sample Size = 150

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.000	0.240	-0.240	0.000	-1.949	0.368	-0.199	0.309	0.200	0.070	0.070	0.000	0.441
2	1.000	0.090	-0.090	0.000	-0.553	0.358	-0.263	0.243	0.200	0.130	-0.130	0.000	0.325
3	1.000	0.370	-0.370	0.000	-2.323	0.548	-0.443	0.324	0.200	0.060	0.060	0.000	0.545
4	1.000	0.270	-0.270	0.000	-2.221	0.413	-0.271	0.312	0.200	0.050	0.050	0.000	0.436
5	1.000	0.210	0.210	0.000	-2.426	0.830	0.514	0.652	0.200	0.020	0.020	0.000	0.384
6	1.000	0.260	-0.260	0.000	-3.053	0.650	-0.523	0.386	0.200	0.020	0.020	0.000	0.419
7	1.000	0.670	-0.670	0.000	-1.871	0.590	-0.491	0.326	0.200	0.060	0.060	0.000	0.790
8	1.000	1.720	-1.720	0.000	-1.991	0.778	-0.741	0.240	0.200	0.070	0.070	0.000	1.231
9	1.000	2.650	-2.650	0.000	-2.896	1.494	-1.446	0.375	0.200	0.110	-0.110	0.000	1.185
10	1.000	0.640	-0.640	0.000	-2.997	0.856	-0.787	0.336	0.200	0.030	0.030	0.000	0.637
11	1.000	0.850	-0.850	0.000	-2.042	0.693	-0.632	0.285	0.200	0.030	0.030	0.000	0.815
12	1.000	0.710	-0.710	0.000	-3.229	1.195	-1.099	0.469	0.200	0.090	-0.090	0.000	0.628
13	1.000	0.510	-0.510	0.000	-2.646	0.745	-0.666	0.333	0.200	0.010	0.010	0.000	0.590
14	1.000	0.780	-0.780	0.000	-1.656	0.704	-0.656	0.255	0.200	0.110	-0.110	0.000	0.474
15	1.000	0.230	0.230	0.000	-1.622	0.566	0.498	0.269	0.200	0.050	0.050	0.000	0.333
16	1.000	0.180	-0.180	0.000	-0.894	0.231	0.036	0.228	0.200	0.050	0.050	0.000	0.321
17	1.000	0.980	-0.980	0.000	-3.783	1.436	-1.343	0.508	0.200	0.020	-0.020	0.000	0.713
18	1.000	0.800	-0.800	0.000	-2.207	0.748	-0.667	0.338	0.200	0.030	0.030	0.000	0.800
19	1.000	0.270	-0.270	0.000	-0.912	0.441	-0.332	0.290	0.200	0.110	-0.110	0.000	0.283
20	1.000	0.740	0.740	0.000	-0.976	2.858	2.844	0.283	0.200	0.000	0.000	0.000	1.202
21	1.000	0.280	-0.280	0.000	-0.248	0.280	0.192	0.204	0.200	0.080	0.080	0.000	0.455
22	1.000	0.690	-0.690	0.000	-1.535	0.442	-0.305	0.320	0.200	0.080	0.080	0.000	0.747
23	1.000	0.390	-0.390	0.000	0.321	0.267	0.061	0.260	0.200	0.040	-0.040	0.000	0.358
24	1.000	0.030	0.030	0.000	-1.732	0.331	0.138	0.300	0.200	0.040	0.040	0.000	0.235
25	1.000	1.220	-1.220	0.000	-2.703	1.082	-1.033	0.322	0.200	0.010	0.010	0.000	1.054
26	1.000	0.010	0.010	0.000	-0.878	0.250	0.002	0.250	0.200	0.020	-0.020	0.000	0.215
27	1.000	0.660	-0.660	0.000	-1.374	0.380	-0.294	0.240	0.200	0.040	0.040	0.000	0.641
28	1.000	0.100	0.100	0.000	0.858	0.484	0.398	0.274	0.200	0.060	0.060	0.000	0.319
29	1.000	1.560	-1.560	0.000	-3.251	1.372	-1.291	0.467	0.200	0.060	0.060	0.000	1.205
30	1.000	0.540	-0.540	0.000	-1.204	0.281	-0.134	0.247	0.200	0.110	0.110	0.000	0.715
31	1.000	0.930	-0.930	0.000	-0.799	0.240	-0.079	0.227	0.200	0.140	0.140	0.000	0.978

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.000	0.970	-0.970	0.000	-0.500	0.348	-0.270	0.220	0.200	0.070	-0.070	0.000	0.470
33	1.000	0.680	-0.680	0.000	-0.420	0.253	0.120	0.222	0.200	0.130	0.130	0.000	0.810
34	1.000	1.610	-1.610	0.000	-2.941	1.239	-1.191	0.343	0.200	0.030	0.030	0.000	1.231
35	1.000	1.460	-1.460	0.000	-2.991	1.235	-1.181	0.364	0.200	0.050	0.050	0.000	1.191
36	1.000	1.130	-1.130	0.000	2.670	1.332	1.230	0.510	0.200	0.040	0.040	0.000	1.178
37	1.000	0.490	0.490	0.000	-0.271	0.563	0.509	0.242	0.200	0.010	-0.010	0.000	0.722
38	1.000	0.020	-0.020	0.000	-1.026	0.289	0.034	0.287	0.200	0.010	0.010	0.000	0.273
39	1.000	0.040	0.040	0.000	1.675	1.031	0.075	1.028	0.200	0.030	-0.030	0.000	0.378
40	1.000	0.350	-0.350	0.000	-0.463	0.323	-0.183	0.266	0.200	0.080	-0.080	0.000	0.298
41	1.000	0.150	0.150	0.000	-1.672	0.381	0.278	0.261	0.200	0.020	0.020	0.000	0.289
42	1.000	1.040	-1.040	0.000	-1.994	0.839	-0.754	0.368	0.200	0.030	-0.030	0.000	0.857
43	1.000	0.150	-0.150	0.000	0.351	0.383	-0.269	0.273	0.200	0.140	-0.140	0.000	0.480
44	1.000	0.400	0.400	0.000	-1.305	0.936	0.725	0.592	0.200	0.000	0.000	0.000	0.661
45	1.000	1.800	-1.800	0.000	-2.726	1.353	-1.296	0.386	0.200	0.090	-0.090	0.000	1.083
46	1.000	0.700	-0.700	0.000	0.253	0.424	0.333	0.264	0.200	0.120	0.120	0.000	0.836
47	1.000	0.200	0.200	0.000	2.418	0.730	0.578	0.446	0.200	0.080	0.080	0.000	0.459
48	1.000	0.100	0.100	0.000	-0.184	0.466	0.386	0.261	0.200	0.090	0.090	0.000	0.265
49	1.000	0.210	0.210	0.000	0.041	0.440	0.371	0.236	0.200	0.070	0.070	0.000	0.254
50	1.000	0.690	0.690	0.000	0.797	0.694	-0.644	0.261	0.200	0.010	-0.010	0.000	0.991

Table E.16. Item-Level Evaluation Criteria Summary Statistics, Condition 3, Sample Size = 150

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.136	0.491	-0.104	0.480	-2.152	1.142	-0.402	1.069	0.200	0.070	0.070	0.000	0.547
2	0.655	0.496	-0.435	0.239	-1.073	1.040	-0.783	0.684	0.200	0.130	-0.130	0.000	0.493
3	1.276	0.503	-0.094	0.494	-2.278	0.894	-0.398	0.800	0.200	0.060	0.060	0.000	0.540
4	1.098	0.520	-0.172	0.491	-2.543	1.275	-0.593	1.129	0.200	0.050	0.050	0.000	0.608
5	0.677	0.359	-0.113	0.341	-3.565	2.826	-0.625	2.756	0.200	0.020	0.020	0.000	0.588
6	1.079	0.598	-0.181	0.570	-2.477	4.976	0.053	4.975	0.200	0.020	0.020	0.000	1.357
7	1.534	0.472	-0.137	0.452	-1.558	0.426	-0.178	0.387	0.200	0.060	0.060	0.000	0.497
8	1.865	0.905	-0.855	0.294	-1.472	0.382	-0.222	0.311	0.200	0.070	0.070	0.000	0.678
9	1.957	1.700	-1.693	0.151	-2.051	0.664	-0.601	0.284	0.200	0.110	-0.110	0.000	0.653
10	1.456	0.562	-0.185	0.531	-2.766	1.226	-0.556	1.093	0.200	0.030	0.030	0.000	0.524
11	1.696	0.427	-0.154	0.398	-1.596	0.430	-0.186	0.387	0.200	0.030	0.030	0.000	0.476
12	1.452	0.579	-0.258	0.519	-2.692	2.187	-0.562	2.114	0.200	0.090	-0.090	0.000	0.740
13	1.413	0.561	-0.097	0.553	-2.529	1.246	-0.549	1.119	0.200	0.010	0.010	0.000	0.564
14	1.285	0.668	-0.495	0.448	-1.591	0.789	-0.591	0.523	0.200	0.110	-0.110	0.000	0.522
15	0.641	0.281	-0.129	0.249	-2.753	1.332	-0.633	1.172	0.200	0.050	0.050	0.000	0.440
16	0.941	0.333	-0.239	0.231	-1.032	0.330	-0.102	0.314	0.200	0.050	0.050	0.000	0.446
17	1.602	0.665	-0.378	0.548	-2.413	5.554	0.028	5.554	0.200	0.020	-0.020	0.000	0.619
18	1.624	0.452	-0.176	0.417	-1.781	0.549	-0.241	0.494	0.200	0.030	0.030	0.000	0.466
19	0.786	0.560	-0.484	0.282	-1.324	0.973	-0.744	0.627	0.200	0.110	-0.110	0.000	0.541
20	0.262	0.123	0.002	0.123	-3.486	4.031	0.334	4.017	0.200	0.000	0.000	0.000	0.657
21	0.922	0.408	-0.358	0.196	-0.340	0.263	0.100	0.243	0.200	0.080	0.080	0.000	0.599
22	1.467	0.483	-0.223	0.428	-1.313	0.403	-0.083	0.394	0.200	0.080	0.080	0.000	0.542
23	0.793	0.627	-0.598	0.189	0.305	0.331	0.045	0.328	0.200	0.040	-0.040	0.000	0.568
24	0.848	0.363	-0.122	0.342	-2.324	1.112	-0.454	1.015	0.200	0.040	0.040	0.000	0.495
25	1.760	0.612	-0.460	0.403	-2.101	0.733	-0.431	0.593	0.200	0.010	0.010	0.000	0.650
26	0.760	0.329	-0.230	0.234	-1.276	0.653	-0.396	0.520	0.200	0.020	-0.020	0.000	0.491
27	1.302	0.576	-0.358	0.451	-1.304	0.464	-0.224	0.407	0.200	0.040	0.040	0.000	0.605
28	0.633	0.307	-0.267	0.151	1.097	0.800	0.637	0.484	0.200	0.060	0.060	0.000	0.631
29	1.916	0.690	-0.644	0.248	-2.360	0.676	-0.400	0.545	0.200	0.060	0.060	0.000	0.681
30	1.369	0.442	-0.171	0.408	-1.074	0.327	-0.004	0.327	0.200	0.110	0.110	0.000	0.587
31	1.566	0.539	-0.364	0.397	-0.650	0.213	0.070	0.201	0.200	0.140	0.140	0.000	0.689

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.188	0.832	-0.782	0.283	-0.470	0.322	-0.240	0.214	0.200	0.070	-0.070	0.000	0.430
33	1.264	0.524	-0.416	0.318	-0.396	0.266	0.144	0.223	0.200	0.130	0.130	0.000	0.688
34	1.903	0.751	-0.707	0.254	-2.122	0.471	-0.372	0.288	0.200	0.030	0.030	0.000	0.657
35	1.897	0.637	-0.563	0.297	-2.204	0.659	-0.394	0.528	0.200	0.050	0.050	0.000	0.648
36	0.684	1.474	-1.446	0.287	4.040	3.063	2.600	1.619	0.200	0.040	0.040	0.000	1.499
37	0.414	0.174	-0.096	0.145	-1.097	0.877	-0.317	0.818	0.200	0.010	-0.010	0.000	0.388
38	0.801	0.336	-0.219	0.255	-1.391	0.607	-0.331	0.509	0.200	0.010	0.010	0.000	0.515
39	0.567	0.425	-0.393	0.160	2.492	1.653	0.892	1.392	0.200	0.030	-0.030	0.000	0.592
40	0.868	0.523	-0.482	0.203	-0.610	0.480	-0.330	0.349	0.200	0.080	-0.080	0.000	0.439
41	0.668	0.294	-0.182	0.230	-2.658	1.288	-0.708	1.076	0.200	0.020	0.020	0.000	0.453
42	1.539	0.680	-0.501	0.461	-1.704	0.802	-0.464	0.654	0.200	0.030	-0.030	0.000	0.656
43	0.617	0.557	-0.533	0.161	0.342	0.505	-0.278	0.421	0.200	0.140	-0.140	0.000	0.536
44	0.446	0.236	-0.154	0.179	-2.780	2.598	-0.750	2.487	0.200	0.000	0.000	0.000	0.658
45	1.854	1.002	-0.946	0.329	-2.044	0.886	-0.614	0.638	0.200	0.090	-0.090	0.000	0.633
46	1.139	0.613	-0.561	0.249	0.229	0.402	0.309	0.258	0.200	0.120	0.120	0.000	0.806
47	0.595	0.268	-0.205	0.174	4.115	4.022	2.275	3.318	0.200	0.080	0.080	0.000	0.736
48	0.638	0.314	-0.262	0.172	-0.446	0.406	0.124	0.387	0.200	0.090	0.090	0.000	0.606
49	0.696	0.184	-0.094	0.158	-0.078	0.406	0.252	0.319	0.200	0.070	0.070	0.000	0.403
50	0.347	0.132	0.037	0.127	1.559	1.152	0.119	1.146	0.200	0.010	-0.010	0.000	0.334

Table E.17. Item-Level Evaluation Criteria Summary Statistics, Condition 4, Sample Size = 150

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.137	0.490	-0.103	0.479	-1.516	0.645	0.234	0.601	0.200	0.070	0.070	0.000	0.512
2	0.656	0.496	-0.435	0.239	-0.718	0.584	-0.428	0.398	0.200	0.130	-0.130	0.000	0.471
3	1.276	0.503	-0.094	0.494	-1.614	0.520	0.267	0.447	0.200	0.060	0.060	0.000	0.526
4	1.098	0.520	-0.172	0.491	-1.753	0.662	0.197	0.632	0.200	0.050	0.050	0.000	0.553
5	0.677	0.359	-0.113	0.341	-2.588	1.183	0.352	1.130	0.200	0.020	0.020	0.000	0.672
6	1.078	0.598	-0.182	0.570	-2.499	1.288	0.031	1.287	0.200	0.020	0.020	0.000	0.689
7	1.534	0.472	-0.136	0.452	-1.159	0.330	0.221	0.244	0.200	0.060	0.060	0.000	0.413
8	1.865	0.905	-0.855	0.295	-1.087	0.267	0.163	0.211	0.200	0.070	0.070	0.000	0.516
9	1.957	1.700	-1.693	0.151	-1.444	0.212	0.006	0.212	0.200	0.110	-0.110	0.000	0.427
10	1.456	0.562	-0.185	0.531	-1.904	0.679	0.306	0.607	0.200	0.030	0.030	0.000	0.650
11	1.696	0.426	-0.154	0.398	-1.184	0.331	0.226	0.243	0.200	0.030	0.030	0.000	0.435
12	1.452	0.580	-0.258	0.519	-2.046	0.862	0.084	0.858	0.200	0.090	-0.090	0.000	0.670
13	1.414	0.561	-0.096	0.553	-1.752	0.666	0.228	0.625	0.200	0.010	0.010	0.000	0.623
14	1.285	0.668	-0.495	0.448	-1.168	0.341	-0.168	0.297	0.200	0.110	-0.110	0.000	0.446
15	0.640	0.281	-0.130	0.249	-1.825	0.736	0.295	0.674	0.200	0.050	0.050	0.000	0.485
16	0.940	0.333	-0.240	0.231	-0.786	0.236	0.144	0.188	0.200	0.050	0.050	0.000	0.388
17	1.602	0.665	-0.378	0.548	-2.273	1.342	0.167	1.332	0.200	0.020	-0.020	0.000	0.719
18	1.624	0.452	-0.176	0.416	-1.315	0.373	0.225	0.298	0.200	0.030	0.030	0.000	0.456
19	0.786	0.560	-0.484	0.282	-0.927	0.505	-0.347	0.367	0.200	0.110	-0.110	0.000	0.481
20	0.262	0.123	0.002	0.123	-2.261	1.956	1.560	1.180	0.200	0.000	0.000	0.000	0.639
21	0.922	0.408	-0.358	0.196	-0.320	0.193	0.120	0.151	0.200	0.080	0.080	0.000	0.573
22	1.467	0.482	-0.223	0.428	-1.011	0.327	0.219	0.243	0.200	0.080	0.080	0.000	0.473
23	0.792	0.627	-0.598	0.189	0.155	0.261	-0.105	0.239	0.200	0.040	-0.040	0.000	0.524
24	0.849	0.363	-0.122	0.342	-1.589	0.644	0.281	0.580	0.200	0.040	0.040	0.000	0.497
25	1.760	0.612	-0.460	0.403	-1.493	0.405	0.177	0.364	0.200	0.010	0.010	0.000	0.510
26	0.760	0.329	-0.230	0.235	-0.892	0.309	-0.012	0.309	0.200	0.020	-0.020	0.000	0.428
27	1.301	0.576	-0.359	0.451	-0.976	0.256	0.104	0.234	0.200	0.040	0.040	0.000	0.508
28	0.633	0.306	-0.267	0.150	0.636	0.381	0.176	0.338	0.200	0.060	0.060	0.000	0.582
29	1.916	0.690	-0.644	0.248	-1.644	0.473	0.316	0.352	0.200	0.060	0.060	0.000	0.574
30	1.369	0.442	-0.172	0.407	-0.848	0.301	0.222	0.203	0.200	0.110	0.110	0.000	0.499
31	1.566	0.539	-0.364	0.397	-0.573	0.195	0.148	0.128	0.200	0.140	0.140	0.000	0.650

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.188	0.831	-0.782	0.282	-0.412	0.236	-0.182	0.150	0.200	0.070	-0.070	0.000	0.404
33	1.264	0.524	-0.416	0.318	-0.397	0.207	0.143	0.149	0.200	0.130	0.130	0.000	0.670
34	1.903	0.751	-0.707	0.254	-1.499	0.321	0.251	0.200	0.200	0.030	0.030	0.000	0.462
35	1.897	0.637	-0.563	0.297	-1.540	0.423	0.270	0.325	0.200	0.050	0.050	0.000	0.531
36	0.683	1.475	-1.447	0.286	2.271	1.203	0.831	0.870	0.200	0.040	0.040	0.000	1.126
37	0.414	0.174	-0.096	0.145	-0.682	0.479	0.099	0.469	0.200	0.010	-0.010	0.000	0.385
38	0.801	0.337	-0.219	0.256	-0.984	0.311	0.076	0.302	0.200	0.010	0.010	0.000	0.437
39	0.567	0.425	-0.393	0.160	1.700	1.236	0.100	1.232	0.200	0.030	-0.030	0.000	0.521
40	0.868	0.524	-0.483	0.203	-0.478	0.297	-0.198	0.222	0.200	0.080	-0.080	0.000	0.410
41	0.668	0.293	-0.183	0.230	-1.766	0.645	0.184	0.618	0.200	0.020	0.020	0.000	0.471
42	1.540	0.680	-0.500	0.460	-1.246	0.383	-0.006	0.383	0.200	0.030	-0.030	0.000	0.477
43	0.617	0.557	-0.533	0.161	0.228	0.502	-0.392	0.313	0.200	0.140	-0.140	0.000	0.465
44	0.446	0.236	-0.154	0.179	-2.000	0.998	0.030	0.997	0.200	0.000	0.000	0.000	0.520
45	1.854	1.001	-0.946	0.329	-1.447	0.387	-0.017	0.387	0.200	0.090	-0.090	0.000	0.424
46	1.139	0.613	-0.561	0.248	0.027	0.208	0.107	0.178	0.200	0.120	0.120	0.000	0.801
47	0.596	0.268	-0.204	0.174	2.037	0.732	0.197	0.705	0.200	0.080	0.080	0.000	0.790
48	0.638	0.313	-0.262	0.172	-0.354	0.329	0.216	0.248	0.200	0.090	0.090	0.000	0.572
49	0.696	0.184	-0.094	0.158	-0.124	0.295	0.206	0.212	0.200	0.070	0.070	0.000	0.371
50	0.347	0.132	0.037	0.127	1.349	1.377	-0.091	1.374	0.200	0.010	-0.010	0.000	0.354

Table E.18. Item-Level Evaluation Criteria Summary Statistics, Condition 5, Sample Size = 150

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.033	0.364	-0.207	0.299	-2.035	0.846	-0.285	0.797	0.195	0.065	0.065	0.007	0.480
2	0.800	0.436	-0.290	0.326	-0.812	0.757	-0.522	0.549	0.203	0.128	-0.127	0.014	0.477
3	1.137	0.402	-0.233	0.327	-2.200	0.661	-0.320	0.578	0.196	0.056	0.056	0.007	0.489
4	1.066	0.470	-0.204	0.424	-2.358	0.998	-0.408	0.911	0.197	0.047	0.047	0.007	0.560
5	0.683	0.284	-0.107	0.263	-3.571	1.484	-0.631	1.343	0.199	0.019	0.019	0.003	0.383
6	0.965	0.479	-0.295	0.377	-3.324	2.220	-0.794	2.073	0.199	0.020	0.019	0.004	0.651
7	1.453	0.487	-0.217	0.436	-1.490	0.386	-0.110	0.370	0.193	0.055	0.053	0.013	0.478
8	2.224	0.846	-0.496	0.685	-1.295	0.305	-0.045	0.302	0.186	0.060	0.056	0.020	0.565
9	2.111	1.619	-1.539	0.503	-1.927	0.646	-0.477	0.436	0.198	0.113	-0.112	0.013	0.591
10	1.207	0.561	-0.433	0.357	-2.775	1.029	-0.565	0.860	0.199	0.029	0.029	0.006	0.520
11	1.606	0.505	-0.244	0.442	-1.530	0.368	-0.120	0.348	0.194	0.027	0.024	0.013	0.482
12	1.235	0.614	-0.475	0.389	-2.773	1.755	-0.643	1.633	0.201	0.090	-0.090	0.005	0.725
13	1.222	0.475	-0.288	0.378	-2.456	1.015	-0.476	0.896	0.198	0.010	0.008	0.006	0.514
14	1.283	0.607	-0.497	0.349	-1.422	0.585	-0.422	0.405	0.200	0.112	-0.111	0.016	0.442
15	0.642	0.261	-0.128	0.228	-2.616	1.215	-0.496	1.109	0.198	0.048	0.048	0.006	0.415
16	1.059	0.258	-0.121	0.228	-0.870	0.282	0.060	0.276	0.189	0.041	0.039	0.012	0.348
17	1.259	0.822	-0.721	0.394	-3.046	2.432	-0.606	2.355	0.200	0.020	-0.020	0.003	0.658
18	1.518	0.459	-0.282	0.361	-1.694	0.423	-0.154	0.394	0.196	0.029	0.026	0.014	0.461
19	0.931	0.459	-0.339	0.310	-1.046	0.685	-0.466	0.503	0.200	0.111	-0.110	0.013	0.470
20	0.298	0.128	0.038	0.122	-3.424	2.262	0.396	2.227	0.200	0.002	0.000	0.002	0.395
21	1.249	0.323	-0.031	0.321	-0.208	0.306	0.232	0.199	0.183	0.066	0.063	0.019	0.412
22	1.437	0.445	-0.253	0.366	-1.221	0.366	0.009	0.366	0.188	0.070	0.068	0.016	0.496
23	1.121	0.492	-0.269	0.413	0.344	0.284	0.084	0.271	0.196	0.051	-0.044	0.024	0.463
24	0.848	0.320	-0.122	0.295	-2.152	0.863	-0.282	0.816	0.197	0.038	0.037	0.008	0.465
25	1.667	0.741	-0.553	0.492	-2.039	0.636	-0.369	0.517	0.198	0.012	0.008	0.009	0.658
26	0.901	0.280	-0.089	0.265	-1.011	0.448	-0.131	0.428	0.200	0.025	-0.020	0.015	0.393
27	1.499	0.471	-0.161	0.443	-1.064	0.323	0.016	0.322	0.194	0.039	0.034	0.020	0.461
28	0.884	0.340	-0.016	0.340	0.918	0.638	0.458	0.445	0.179	0.043	0.039	0.019	0.518
29	1.816	0.875	-0.744	0.459	-2.358	0.778	-0.398	0.668	0.195	0.055	0.055	0.008	0.685
30	1.404	0.390	-0.136	0.366	-0.964	0.301	0.106	0.282	0.184	0.096	0.094	0.016	0.494
31	1.833	0.515	-0.097	0.506	-0.547	0.254	0.173	0.186	0.164	0.105	0.104	0.017	0.521

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.507	0.621	-0.463	0.414	-0.302	0.223	-0.072	0.211	0.200	0.074	-0.070	0.025	0.346
33	1.537	0.374	-0.143	0.346	-0.312	0.305	0.228	0.203	0.165	0.096	0.095	0.017	0.488
34	1.850	0.874	-0.760	0.431	-2.059	0.472	-0.309	0.357	0.195	0.027	0.025	0.009	0.648
35	1.809	0.807	-0.651	0.478	-2.161	0.639	-0.351	0.533	0.193	0.044	0.043	0.011	0.673
36	1.357	0.905	-0.773	0.471	2.192	1.179	0.752	0.908	0.154	0.025	-0.007	0.025	0.890
37	0.465	0.168	-0.045	0.162	-0.862	0.600	-0.082	0.595	0.199	0.014	-0.012	0.008	0.363
38	0.906	0.301	-0.114	0.278	-1.173	0.481	-0.113	0.467	0.199	0.019	0.009	0.016	0.449
39	0.686	0.417	-0.274	0.315	2.264	1.306	0.664	1.125	0.187	0.047	-0.043	0.020	0.582
40	1.011	0.427	-0.339	0.260	-0.462	0.357	-0.182	0.307	0.197	0.084	-0.083	0.015	0.413
41	0.717	0.256	-0.133	0.219	-2.354	0.953	-0.404	0.863	0.199	0.020	0.019	0.005	0.392
42	1.561	0.645	-0.479	0.433	-1.530	0.561	-0.290	0.481	0.197	0.037	-0.033	0.016	0.595
43	0.753	0.484	-0.397	0.278	0.399	0.451	-0.221	0.394	0.200	0.141	-0.140	0.015	0.558
44	0.508	0.226	-0.093	0.206	-2.668	1.399	-0.638	1.245	0.200	0.004	-0.001	0.004	0.418
45	1.863	1.058	-0.937	0.491	-1.932	0.756	-0.502	0.565	0.199	0.091	-0.091	0.011	0.601
46	1.695	0.379	-0.005	0.379	0.208	0.347	0.288	0.193	0.157	0.080	0.077	0.019	0.486
47	0.745	0.321	-0.055	0.317	3.075	1.968	1.235	1.532	0.165	0.050	0.045	0.021	0.623
48	0.797	0.271	-0.103	0.250	-0.320	0.425	0.250	0.344	0.191	0.081	0.081	0.010	0.481
49	0.798	0.231	0.008	0.231	-0.028	0.412	0.302	0.280	0.188	0.060	0.058	0.014	0.380
50	0.353	0.158	0.043	0.152	1.616	1.090	0.176	1.076	0.203	0.013	-0.008	0.010	0.358

Table E.19. Item-Level Evaluation Criteria Summary Statistics, Condition 0, Sample Size = 100

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.323	0.488	0.083	0.480	-1.776	0.650	-0.026	0.649	0.197	0.066	0.067	0.014	0.480
2	0.964	0.479	-0.126	0.462	-0.741	0.718	-0.451	0.559	0.207	0.127	-0.123	0.031	0.504
3	1.533	0.654	0.163	0.634	-1.989	0.738	-0.109	0.730	0.200	0.059	0.060	0.013	0.552
4	1.466	0.790	0.196	0.766	-2.074	0.985	-0.124	0.977	0.200	0.049	0.050	0.011	0.588
5	0.854	0.391	0.064	0.385	-3.229	1.486	-0.289	1.457	0.202	0.020	0.022	0.008	0.468
6	1.268	0.705	0.008	0.705	-3.042	2.168	-0.512	2.106	0.201	0.020	0.021	0.008	0.708
7	1.990	0.888	0.320	0.828	-1.304	0.409	0.076	0.402	0.199	0.059	0.059	0.006	0.433
8	2.948	1.166	0.228	1.143	-1.152	0.333	0.099	0.318	0.196	0.068	0.066	0.015	0.538
9	3.049	1.070	-0.601	0.885	-1.633	0.386	-0.183	0.340	0.206	0.108	-0.104	0.030	0.500
10	1.601	0.632	-0.039	0.631	-2.559	1.280	-0.349	1.231	0.202	0.030	0.032	0.009	0.554
11	2.167	0.822	0.317	0.758	-1.357	0.437	0.053	0.434	0.197	0.030	0.027	0.012	0.513
12	1.819	0.764	0.109	0.756	-2.614	1.409	-0.484	1.323	0.204	0.089	-0.086	0.022	0.596
13	1.642	0.650	0.132	0.637	-2.159	0.998	-0.179	0.982	0.202	0.014	0.012	0.006	0.575
14	1.747	0.848	-0.033	0.847	-1.291	0.579	-0.291	0.501	0.207	0.108	-0.103	0.032	0.512
15	0.821	0.424	0.051	0.421	-2.283	1.062	-0.163	1.050	0.200	0.049	0.050	0.012	0.463
16	1.268	0.391	0.088	0.381	-0.824	0.367	0.106	0.351	0.194	0.044	0.044	0.002	0.387
17	1.933	0.785	-0.047	0.784	-2.782	2.025	-0.342	1.996	0.204	0.020	-0.016	0.011	0.637
18	2.056	0.773	0.256	0.729	-1.517	0.469	0.023	0.468	0.202	0.034	0.032	0.012	0.453
19	1.158	0.560	-0.112	0.548	-1.008	0.733	-0.428	0.596	0.205	0.109	-0.106	0.028	0.527
20	0.376	0.199	0.116	0.162	-3.015	2.253	0.805	2.105	0.203	0.003	0.003	0.001	0.452
21	1.519	0.704	0.239	0.662	-0.222	0.328	0.218	0.245	0.192	0.072	0.072	0.008	0.466
22	1.864	0.644	0.174	0.620	-1.089	0.409	0.141	0.384	0.195	0.075	0.075	0.007	0.464
23	1.292	0.586	-0.098	0.578	0.297	0.338	0.037	0.336	0.201	0.048	-0.039	0.029	0.464
24	1.081	0.502	0.111	0.490	-1.941	1.045	-0.071	1.042	0.201	0.040	0.041	0.011	0.546
25	2.346	0.857	0.126	0.848	-1.767	0.603	-0.097	0.595	0.204	0.018	0.014	0.012	0.576
26	1.069	0.498	0.079	0.491	-0.973	0.490	-0.093	0.481	0.203	0.023	-0.017	0.016	0.462
27	1.851	0.808	0.191	0.785	-1.006	0.369	0.074	0.361	0.200	0.043	0.040	0.015	0.486
28	1.032	0.479	0.132	0.460	0.861	0.641	0.401	0.500	0.187	0.049	0.047	0.011	0.536
29	2.651	0.692	0.091	0.686	-1.974	0.587	-0.014	0.587	0.198	0.058	0.058	0.010	0.578
30	1.789	0.846	0.249	0.809	-0.851	0.351	0.219	0.275	0.193	0.103	0.103	0.006	0.446
31	2.217	0.808	0.287	0.756	-0.522	0.303	0.198	0.230	0.177	0.117	0.117	0.006	0.514

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.980	0.853	0.010	0.853	-0.280	0.245	-0.050	0.240	0.206	0.072	-0.064	0.033	0.397
33	1.858	0.622	0.178	0.596	-0.305	0.318	0.235	0.214	0.179	0.109	0.109	0.004	0.474
34	2.766	0.811	0.156	0.796	-1.744	0.426	0.007	0.426	0.200	0.030	0.030	0.003	0.464
35	2.613	0.819	0.153	0.804	-1.836	0.454	-0.026	0.454	0.200	0.049	0.050	0.007	0.555
36	1.690	0.788	-0.440	0.653	1.951	0.884	0.511	0.721	0.164	0.025	0.004	0.025	0.744
37	0.553	0.250	0.043	0.247	-0.817	0.680	-0.037	0.679	0.203	0.012	-0.007	0.010	0.447
38	1.062	0.380	0.042	0.378	-1.090	0.567	-0.030	0.566	0.202	0.016	0.012	0.011	0.489
39	0.890	0.457	-0.070	0.452	2.101	1.293	0.501	1.192	0.198	0.040	-0.032	0.024	0.615
40	1.218	0.398	-0.132	0.375	-0.419	0.359	-0.139	0.331	0.202	0.082	-0.079	0.024	0.419
41	0.879	0.317	0.029	0.316	-2.121	0.917	-0.171	0.901	0.202	0.021	0.022	0.008	0.431
42	2.059	0.892	0.019	0.892	-1.375	0.488	-0.135	0.469	0.203	0.035	-0.027	0.023	0.548
43	0.965	0.564	-0.185	0.533	0.360	0.537	-0.260	0.470	0.208	0.137	-0.132	0.036	0.588
44	0.639	0.303	0.039	0.301	-2.424	1.439	-0.394	1.384	0.202	0.005	0.002	0.004	0.468
45	2.746	0.913	-0.054	0.911	-1.671	0.725	-0.241	0.684	0.204	0.089	-0.086	0.025	0.538
46	2.021	0.658	0.321	0.574	0.176	0.328	0.256	0.205	0.170	0.090	0.090	0.008	0.480
47	0.981	0.419	0.181	0.378	2.515	1.281	0.675	1.089	0.180	0.062	0.060	0.016	0.600
48	0.948	0.374	0.048	0.371	-0.323	0.532	0.247	0.471	0.195	0.085	0.085	0.012	0.525
49	0.933	0.402	0.143	0.375	-0.011	0.498	0.319	0.382	0.195	0.065	0.065	0.006	0.437
50	0.440	0.231	0.130	0.191	1.399	1.097	-0.041	1.096	0.208	0.010	-0.002	0.010	0.474

Table E.20. Item-Level Evaluation Criteria Summary Statistics, Condition 1, Sample Size = 100

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.198	0.526	-0.042	0.524	-2.160	1.239	-0.410	1.169	0.198	0.067	0.068	0.014	0.572
2	0.668	0.517	-0.422	0.299	-1.114	1.228	-0.824	0.910	0.197	0.136	-0.133	0.029	0.526
3	1.345	0.578	-0.025	0.578	-2.422	1.317	-0.542	1.200	0.200	0.059	0.060	0.013	0.602
4	1.202	0.548	-0.068	0.544	-2.258	2.304	-0.308	2.284	0.200	0.048	0.050	0.012	0.738
5	0.703	0.404	-0.087	0.395	-2.623	4.863	0.317	4.852	0.202	0.020	0.022	0.009	1.076
6	1.065	0.662	-0.195	0.633	-1.522	6.740	1.008	6.664	0.201	0.020	0.021	0.008	1.738
7	1.595	0.470	-0.075	0.464	-1.543	0.539	-0.163	0.513	0.196	0.056	0.056	0.004	0.453
8	1.853	0.956	-0.867	0.403	-1.481	0.447	-0.231	0.383	0.188	0.058	0.058	0.001	0.672
9	1.956	1.737	-1.695	0.381	-2.045	0.671	-0.595	0.311	0.198	0.115	-0.112	0.027	0.642
10	1.411	0.654	-0.229	0.613	-2.268	3.533	-0.058	3.533	0.202	0.030	0.032	0.010	0.991
11	1.682	0.478	-0.169	0.447	-1.642	0.622	-0.232	0.577	0.195	0.026	0.025	0.008	0.539
12	1.493	0.645	-0.217	0.608	-1.863	4.277	0.267	4.268	0.203	0.090	-0.087	0.022	1.172
13	1.422	0.578	-0.088	0.572	-2.401	2.215	-0.421	2.175	0.202	0.012	0.012	0.001	0.582
14	1.322	0.729	-0.458	0.566	-1.677	1.047	-0.677	0.799	0.201	0.113	-0.109	0.027	0.569
15	0.706	0.393	-0.064	0.388	-2.379	2.565	-0.259	2.552	0.200	0.049	0.050	0.013	0.604
16	0.992	0.395	-0.188	0.347	-1.059	0.482	-0.129	0.464	0.190	0.039	0.040	0.007	0.465
17	1.543	0.783	-0.437	0.650	-2.097	6.489	0.343	6.479	0.202	0.021	-0.018	0.010	1.078
18	1.637	0.526	-0.163	0.500	-1.896	1.052	-0.356	0.990	0.198	0.029	0.028	0.005	0.517
19	0.832	0.593	-0.439	0.399	-1.409	1.225	-0.829	0.902	0.196	0.117	-0.114	0.027	0.579
20	0.276	0.143	0.016	0.142	-1.984	5.864	1.836	5.569	0.202	0.002	0.002	0.001	0.989
21	0.970	0.455	-0.310	0.333	-0.378	0.323	0.062	0.318	0.180	0.060	0.060	0.006	0.620
22	1.523	0.526	-0.167	0.499	-1.330	0.551	-0.100	0.542	0.192	0.072	0.072	0.010	0.541
23	0.801	0.649	-0.589	0.272	0.298	0.441	0.038	0.440	0.186	0.059	-0.055	0.022	0.619
24	0.934	0.464	-0.036	0.462	-1.853	2.851	0.017	2.851	0.200	0.038	0.040	0.011	0.624
25	1.741	0.659	-0.479	0.452	-2.152	0.975	-0.482	0.848	0.200	0.013	0.010	0.009	0.649
26	0.772	0.397	-0.218	0.332	-1.337	0.803	-0.457	0.661	0.196	0.028	-0.024	0.013	0.543
27	1.313	0.594	-0.347	0.482	-1.342	0.582	-0.262	0.520	0.192	0.032	0.032	0.005	0.627
28	0.657	0.312	-0.243	0.195	1.057	0.856	0.597	0.614	0.178	0.039	0.038	0.008	0.631
29	1.931	0.714	-0.629	0.339	-2.375	0.907	-0.415	0.806	0.196	0.055	0.056	0.011	0.700
30	1.408	0.470	-0.132	0.451	-1.046	0.363	0.024	0.362	0.188	0.097	0.098	0.013	0.567
31	1.570	0.574	-0.360	0.447	-0.680	0.283	0.040	0.280	0.172	0.112	0.112	0.010	0.650

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.246	0.816	-0.724	0.376	-0.452	0.345	-0.222	0.263	0.188	0.085	-0.082	0.024	0.472
33	1.321	0.545	-0.360	0.409	-0.424	0.273	0.116	0.247	0.172	0.101	0.102	0.009	0.630
34	1.923	0.769	-0.687	0.346	-2.155	0.769	-0.405	0.654	0.196	0.026	0.026	0.004	0.642
35	1.892	0.675	-0.569	0.365	-2.221	0.708	-0.411	0.577	0.197	0.046	0.047	0.010	0.648
36	0.752	1.432	-1.378	0.392	3.487	2.622	2.047	1.639	0.156	0.022	-0.004	0.022	1.353
37	0.413	0.211	-0.097	0.187	-1.216	1.186	-0.436	1.103	0.199	0.014	-0.011	0.009	0.485
38	0.823	0.373	-0.197	0.317	-1.404	0.780	-0.344	0.700	0.197	0.009	0.007	0.006	0.572
39	0.585	0.442	-0.375	0.234	2.815	2.360	1.215	2.023	0.188	0.048	-0.042	0.023	0.675
40	0.914	0.537	-0.436	0.312	-0.595	0.523	-0.315	0.417	0.192	0.091	-0.088	0.023	0.468
41	0.706	0.337	-0.144	0.305	-2.550	1.916	-0.600	1.820	0.201	0.019	0.021	0.008	0.662
42	1.508	0.762	-0.532	0.546	-1.770	0.904	-0.530	0.733	0.197	0.037	-0.033	0.017	0.704
43	0.627	0.570	-0.523	0.227	0.375	0.641	-0.245	0.592	0.194	0.149	-0.146	0.031	0.606
44	0.485	0.275	-0.115	0.250	-1.860	4.241	0.170	4.238	0.201	0.004	0.001	0.004	0.747
45	1.840	1.063	-0.960	0.455	-1.826	2.129	-0.396	2.092	0.198	0.095	-0.092	0.024	0.678
46	1.199	0.608	-0.501	0.344	0.176	0.378	0.256	0.278	0.159	0.079	0.079	0.003	0.720
47	0.654	0.289	-0.147	0.250	3.474	2.649	1.634	2.086	0.174	0.057	0.054	0.016	0.651
48	0.672	0.336	-0.228	0.246	-0.491	0.627	0.079	0.622	0.189	0.078	0.079	0.013	0.644
49	0.711	0.246	-0.079	0.233	-0.059	0.536	0.271	0.462	0.189	0.058	0.059	0.009	0.482
50	0.340	0.152	0.030	0.149	1.813	1.667	0.373	1.624	0.202	0.013	-0.008	0.010	0.418

Table E.21. Item-Level Evaluation Criteria Summary Statistics, Condition 2, Sample Size = 100

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.000	0.240	-0.240	0.000	-1.950	0.423	-0.200	0.372	0.200	0.070	0.070	0.000	0.473
2	1.000	0.090	-0.090	0.000	-0.536	0.374	-0.246	0.282	0.200	0.130	-0.130	0.000	0.354
3	1.000	0.370	-0.370	0.000	-2.347	0.610	-0.467	0.393	0.200	0.060	0.060	0.000	0.559
4	1.000	0.270	-0.270	0.000	-2.267	0.507	-0.317	0.396	0.200	0.050	0.050	0.000	0.474
5	1.000	0.210	0.210	0.000	-2.531	1.007	0.409	0.920	0.200	0.020	0.020	0.000	0.437
6	1.000	0.260	-0.260	0.000	-2.910	1.528	-0.380	1.480	0.200	0.020	0.020	0.000	0.618
7	1.000	0.670	-0.670	0.000	-1.878	0.635	-0.498	0.393	0.200	0.060	0.060	0.000	0.769
8	1.000	1.720	-1.720	0.000	-1.971	0.777	-0.721	0.287	0.200	0.070	0.070	0.000	1.192
9	1.000	2.650	-2.650	0.000	-2.883	1.474	-1.433	0.346	0.200	0.110	-0.110	0.000	1.140
10	1.000	0.640	-0.640	0.000	-2.996	0.881	-0.786	0.398	0.200	0.030	0.030	0.000	0.627
11	1.000	0.850	-0.850	0.000	-2.031	0.715	-0.621	0.354	0.200	0.030	0.030	0.000	0.803
12	1.000	0.710	-0.710	0.000	-3.252	1.210	-1.122	0.452	0.200	0.090	-0.090	0.000	0.630
13	1.000	0.510	-0.510	0.000	-2.629	0.748	-0.649	0.371	0.200	0.010	0.010	0.000	0.556
14	1.000	0.780	-0.780	0.000	-1.663	0.725	-0.663	0.293	0.200	0.110	-0.110	0.000	0.493
15	1.000	0.230	0.230	0.000	-1.612	0.625	0.508	0.364	0.200	0.050	0.050	0.000	0.407
16	1.000	0.180	-0.180	0.000	-0.877	0.283	0.053	0.278	0.200	0.050	0.050	0.000	0.330
17	1.000	0.980	-0.980	0.000	-3.805	1.485	-1.365	0.586	0.200	0.020	-0.020	0.000	0.691
18	1.000	0.800	-0.800	0.000	-2.230	0.789	-0.690	0.382	0.200	0.030	0.030	0.000	0.782
19	1.000	0.270	-0.270	0.000	-0.920	0.497	-0.340	0.363	0.200	0.110	-0.110	0.000	0.310
20	1.000	0.740	0.740	0.000	-1.015	2.850	2.805	0.506	0.200	0.000	0.000	0.000	1.169
21	1.000	0.280	-0.280	0.000	-0.230	0.343	0.210	0.271	0.200	0.080	0.080	0.000	0.476
22	1.000	0.690	-0.690	0.000	-1.525	0.456	-0.295	0.348	0.200	0.080	0.080	0.000	0.736
23	1.000	0.390	-0.390	0.000	0.358	0.357	0.098	0.343	0.200	0.040	-0.040	0.000	0.400
24	1.000	0.030	0.030	0.000	-1.701	0.455	0.169	0.422	0.200	0.040	0.040	0.000	0.315
25	1.000	1.220	-1.220	0.000	-2.693	1.082	-1.023	0.351	0.200	0.010	0.010	0.000	1.032
26	1.000	0.010	0.010	0.000	-0.848	0.293	0.032	0.291	0.200	0.020	-0.020	0.000	0.245
27	1.000	0.660	-0.660	0.000	-1.376	0.434	-0.296	0.317	0.200	0.040	0.040	0.000	0.648
28	1.000	0.100	0.100	0.000	0.910	0.571	0.450	0.350	0.200	0.060	0.060	0.000	0.387
29	1.000	1.560	-1.560	0.000	-3.214	1.344	-1.254	0.482	0.200	0.060	0.060	0.000	1.163
30	1.000	0.540	-0.540	0.000	-1.162	0.313	-0.092	0.299	0.200	0.110	0.110	0.000	0.687
31	1.000	0.930	-0.930	0.000	-0.782	0.291	-0.062	0.284	0.200	0.140	0.140	0.000	0.958

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.000	0.970	-0.970	0.000	-0.466	0.359	-0.236	0.270	0.200	0.070	-0.070	0.000	0.483
33	1.000	0.680	-0.680	0.000	-0.412	0.318	0.128	0.292	0.200	0.130	0.130	0.000	0.813
34	1.000	1.610	-1.610	0.000	-2.925	1.241	-1.175	0.401	0.200	0.030	0.030	0.000	1.188
35	1.000	1.460	-1.460	0.000	-2.993	1.245	-1.183	0.387	0.200	0.050	0.050	0.000	1.146
36	1.000	1.130	-1.130	0.000	2.733	1.392	1.293	0.513	0.200	0.040	0.040	0.000	1.194
37	1.000	0.490	0.490	0.000	-0.258	0.617	0.522	0.330	0.200	0.010	-0.010	0.000	0.727
38	1.000	0.020	-0.020	0.000	-0.983	0.355	0.077	0.346	0.200	0.010	0.010	0.000	0.321
39	1.000	0.040	0.040	0.000	1.715	0.470	0.115	0.455	0.200	0.030	-0.030	0.000	0.420
40	1.000	0.350	-0.350	0.000	-0.422	0.310	-0.142	0.276	0.200	0.080	-0.080	0.000	0.299
41	1.000	0.150	0.150	0.000	-1.671	0.448	0.279	0.351	0.200	0.020	0.020	0.000	0.331
42	1.000	1.040	-1.040	0.000	-1.977	0.827	-0.737	0.374	0.200	0.030	-0.030	0.000	0.833
43	1.000	0.150	-0.150	0.000	0.385	0.425	-0.235	0.354	0.200	0.140	-0.140	0.000	0.515
44	1.000	0.400	0.400	0.000	-1.375	1.172	0.655	0.972	0.200	0.000	0.000	0.000	0.636
45	1.000	1.800	-1.800	0.000	-2.700	1.327	-1.270	0.383	0.200	0.090	-0.090	0.000	1.028
46	1.000	0.700	-0.700	0.000	0.293	0.489	0.373	0.316	0.200	0.120	0.120	0.000	0.844
47	1.000	0.200	0.200	0.000	2.454	0.829	0.614	0.557	0.200	0.080	0.080	0.000	0.492
48	1.000	0.100	0.100	0.000	-0.158	0.537	0.412	0.344	0.200	0.090	0.090	0.000	0.327
49	1.000	0.210	0.210	0.000	0.095	0.554	0.425	0.355	0.200	0.070	0.070	0.000	0.327
50	1.000	0.690	0.690	0.000	0.841	0.700	-0.599	0.361	0.200	0.010	-0.010	0.000	0.987

Table E.22. Item-Level Evaluation Criteria Summary Statistics, Condition 3, Sample Size = 100

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.198	0.526	-0.042	0.524	-2.160	1.241	-0.410	1.172	0.200	0.070	0.070	0.000	0.571
2	0.668	0.517	-0.422	0.299	-1.105	1.227	-0.815	0.918	0.200	0.130	-0.130	0.000	0.519
3	1.345	0.578	-0.025	0.578	-2.426	1.319	-0.546	1.200	0.200	0.060	0.060	0.000	0.598
4	1.202	0.548	-0.068	0.544	-2.260	2.308	-0.310	2.287	0.200	0.050	0.050	0.000	0.737
5	0.703	0.404	-0.087	0.395	-2.626	4.867	0.314	4.857	0.200	0.020	0.020	0.000	1.076
6	1.065	0.662	-0.195	0.633	-1.523	6.749	1.007	6.674	0.200	0.020	0.020	0.000	1.738
7	1.595	0.470	-0.075	0.464	-1.543	0.541	-0.163	0.516	0.200	0.060	0.060	0.000	0.454
8	1.853	0.956	-0.867	0.403	-1.475	0.446	-0.225	0.385	0.200	0.070	0.070	0.000	0.689
9	1.956	1.737	-1.695	0.381	-2.050	0.675	-0.600	0.309	0.200	0.110	-0.110	0.000	0.650
10	1.411	0.654	-0.229	0.613	-2.273	3.536	-0.063	3.535	0.200	0.030	0.030	0.000	0.991
11	1.682	0.478	-0.169	0.447	-1.641	0.624	-0.231	0.580	0.200	0.030	0.030	0.000	0.537
12	1.493	0.645	-0.217	0.608	-1.870	4.279	0.260	4.271	0.200	0.090	-0.090	0.000	1.172
13	1.422	0.578	-0.088	0.572	-2.408	2.218	-0.428	2.176	0.200	0.010	0.010	0.000	0.581
14	1.322	0.729	-0.458	0.566	-1.679	1.049	-0.679	0.799	0.200	0.110	-0.110	0.000	0.565
15	0.706	0.393	-0.064	0.388	-2.379	2.569	-0.259	2.556	0.200	0.050	0.050	0.000	0.602
16	0.992	0.395	-0.188	0.347	-1.041	0.486	-0.111	0.474	0.200	0.050	0.050	0.000	0.467
17	1.543	0.783	-0.437	0.650	-2.101	6.495	0.339	6.486	0.200	0.020	-0.020	0.000	1.078
18	1.637	0.526	-0.163	0.500	-1.899	1.052	-0.359	0.989	0.200	0.030	0.030	0.000	0.517
19	0.832	0.593	-0.439	0.399	-1.401	1.222	-0.821	0.906	0.200	0.110	-0.110	0.000	0.576
20	0.276	0.143	0.016	0.142	-1.988	5.888	1.832	5.596	0.200	0.000	0.000	0.000	0.991
21	0.970	0.455	-0.310	0.333	-0.333	0.350	0.107	0.333	0.200	0.080	0.080	0.000	0.641
22	1.523	0.526	-0.167	0.499	-1.322	0.554	-0.092	0.546	0.200	0.080	0.080	0.000	0.545
23	0.801	0.649	-0.589	0.272	0.350	0.460	0.090	0.451	0.200	0.040	-0.040	0.000	0.623
24	0.934	0.464	-0.036	0.462	-1.853	2.855	0.017	2.855	0.200	0.040	0.040	0.000	0.621
25	1.741	0.659	-0.479	0.452	-2.158	0.977	-0.488	0.847	0.200	0.010	0.010	0.000	0.653
26	0.772	0.397	-0.218	0.332	-1.327	0.804	-0.447	0.668	0.200	0.020	-0.020	0.000	0.541
27	1.313	0.594	-0.347	0.482	-1.333	0.583	-0.253	0.525	0.200	0.040	0.040	0.000	0.628
28	0.657	0.312	-0.243	0.195	1.165	0.936	0.705	0.617	0.200	0.060	0.060	0.000	0.654
29	1.931	0.714	-0.629	0.339	-2.377	0.909	-0.417	0.808	0.200	0.060	0.060	0.000	0.706
30	1.408	0.470	-0.132	0.451	-1.032	0.371	0.038	0.369	0.200	0.110	0.110	0.000	0.579
31	1.570	0.574	-0.360	0.447	-0.643	0.302	0.077	0.292	0.200	0.140	0.140	0.000	0.705

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.246	0.816	-0.724	0.376	-0.431	0.337	-0.201	0.270	0.200	0.070	-0.070	0.000	0.468
33	1.321	0.545	-0.360	0.409	-0.376	0.307	0.164	0.260	0.200	0.130	0.130	0.000	0.682
34	1.923	0.769	-0.687	0.346	-2.158	0.772	-0.408	0.656	0.200	0.030	0.030	0.000	0.652
35	1.892	0.675	-0.569	0.365	-2.223	0.710	-0.413	0.577	0.200	0.050	0.050	0.000	0.653
36	0.752	1.432	-1.378	0.392	3.954	3.046	2.514	1.720	0.200	0.040	0.040	0.000	1.448
37	0.413	0.211	-0.097	0.187	-1.212	1.196	-0.432	1.116	0.200	0.010	-0.010	0.000	0.481
38	0.823	0.373	-0.197	0.317	-1.395	0.782	-0.335	0.707	0.200	0.010	0.010	0.000	0.570
39	0.585	0.442	-0.375	0.234	2.880	2.309	1.280	1.922	0.200	0.030	-0.030	0.000	0.675
40	0.914	0.537	-0.436	0.312	-0.575	0.519	-0.295	0.427	0.200	0.080	-0.080	0.000	0.459
41	0.706	0.337	-0.144	0.305	-2.550	1.917	-0.600	1.821	0.200	0.020	0.020	0.000	0.661
42	1.508	0.762	-0.532	0.546	-1.770	0.903	-0.530	0.731	0.200	0.030	-0.030	0.000	0.706
43	0.627	0.570	-0.523	0.227	0.398	0.627	-0.222	0.586	0.200	0.140	-0.140	0.000	0.593
44	0.485	0.275	-0.115	0.250	-1.858	4.250	0.172	4.247	0.200	0.000	0.000	0.000	0.745
45	1.840	1.063	-0.960	0.455	-1.829	2.131	-0.399	2.093	0.200	0.090	-0.090	0.000	0.681
46	1.199	0.608	-0.501	0.344	0.267	0.458	0.347	0.299	0.200	0.120	0.120	0.000	0.803
47	0.654	0.289	-0.147	0.250	3.745	2.807	1.905	2.061	0.200	0.080	0.080	0.000	0.703
48	0.672	0.336	-0.228	0.246	-0.452	0.644	0.118	0.633	0.200	0.090	0.090	0.000	0.652
49	0.711	0.246	-0.079	0.233	-0.019	0.569	0.311	0.476	0.200	0.070	0.070	0.000	0.482
50	0.340	0.152	0.030	0.149	1.775	1.599	0.335	1.564	0.200	0.010	-0.010	0.000	0.409

Table E.23. Item-Level Evaluation Criteria Summary Statistics, Condition 4, Sample Size = 100

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.198	0.526	-0.043	0.524	-1.534	0.710	0.216	0.676	0.200	0.070	0.070	0.000	0.547
2	0.668	0.517	-0.422	0.299	-0.762	0.720	-0.472	0.543	0.200	0.130	-0.130	0.000	0.487
3	1.344	0.578	-0.026	0.578	-1.701	0.716	0.179	0.693	0.200	0.060	0.060	0.000	0.603
4	1.202	0.548	-0.068	0.544	-1.715	0.702	0.235	0.661	0.200	0.050	0.050	0.000	0.627
5	0.704	0.404	-0.086	0.395	-2.732	1.535	0.208	1.521	0.200	0.020	0.020	0.000	0.697
6	1.065	0.662	-0.195	0.633	-2.885	2.102	-0.355	2.072	0.200	0.020	0.020	0.000	0.752
7	1.595	0.469	-0.075	0.463	-1.168	0.396	0.212	0.335	0.200	0.060	0.060	0.000	0.417
8	1.854	0.955	-0.866	0.403	-1.104	0.310	0.146	0.273	0.200	0.070	0.070	0.000	0.554
9	1.956	1.737	-1.695	0.381	-1.452	0.256	-0.002	0.256	0.200	0.110	-0.110	0.000	0.482
10	1.411	0.655	-0.229	0.613	-2.203	1.451	0.007	1.451	0.200	0.030	0.030	0.000	0.695
11	1.681	0.478	-0.169	0.447	-1.227	0.412	0.183	0.369	0.200	0.030	0.030	0.000	0.479
12	1.493	0.645	-0.217	0.608	-2.323	1.682	-0.193	1.671	0.200	0.090	-0.090	0.000	0.707
13	1.422	0.578	-0.088	0.572	-1.862	1.019	0.118	1.012	0.200	0.010	0.010	0.000	0.659
14	1.322	0.729	-0.458	0.566	-1.228	0.517	-0.228	0.464	0.200	0.110	-0.110	0.000	0.510
15	0.706	0.393	-0.064	0.387	-1.856	0.982	0.264	0.946	0.200	0.050	0.050	0.000	0.575
16	0.992	0.394	-0.188	0.347	-0.805	0.312	0.125	0.286	0.200	0.050	0.050	0.000	0.415
17	1.543	0.783	-0.437	0.650	-2.575	1.987	-0.135	1.983	0.200	0.020	-0.020	0.000	0.771
18	1.637	0.525	-0.163	0.500	-1.384	0.596	0.156	0.575	0.200	0.030	0.030	0.000	0.501
19	0.831	0.593	-0.439	0.399	-0.994	0.678	-0.414	0.537	0.200	0.110	-0.110	0.000	0.530
20	0.276	0.142	0.016	0.141	-1.993	2.238	1.827	1.292	0.200	0.000	0.000	0.000	0.680
21	0.970	0.455	-0.310	0.333	-0.321	0.242	0.119	0.210	0.200	0.080	0.080	0.000	0.603
22	1.523	0.526	-0.167	0.499	-1.014	0.400	0.216	0.337	0.200	0.080	0.080	0.000	0.491
23	0.801	0.649	-0.589	0.272	0.166	0.334	-0.094	0.320	0.200	0.040	-0.040	0.000	0.564
24	0.934	0.464	-0.036	0.462	-1.608	0.921	0.262	0.883	0.200	0.040	0.040	0.000	0.580
25	1.741	0.658	-0.479	0.452	-1.549	0.521	0.121	0.507	0.200	0.010	0.010	0.000	0.515
26	0.772	0.397	-0.218	0.332	-0.940	0.407	-0.060	0.402	0.200	0.020	-0.020	0.000	0.483
27	1.313	0.594	-0.347	0.482	-1.009	0.325	0.071	0.317	0.200	0.040	0.040	0.000	0.530
28	0.657	0.312	-0.244	0.195	0.682	0.485	0.222	0.431	0.200	0.060	0.060	0.000	0.578
29	1.931	0.714	-0.629	0.339	-1.661	0.583	0.299	0.501	0.200	0.060	0.060	0.000	0.627
30	1.409	0.470	-0.131	0.451	-0.828	0.341	0.243	0.239	0.200	0.110	0.110	0.000	0.515
31	1.570	0.573	-0.360	0.447	-0.568	0.247	0.152	0.194	0.200	0.140	0.140	0.000	0.666

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.246	0.816	-0.724	0.376	-0.407	0.258	-0.177	0.188	0.200	0.070	-0.070	0.000	0.434
33	1.320	0.546	-0.360	0.410	-0.385	0.238	0.155	0.181	0.200	0.130	0.130	0.000	0.662
34	1.923	0.769	-0.687	0.346	-1.529	0.464	0.221	0.408	0.200	0.030	0.030	0.000	0.496
35	1.891	0.676	-0.569	0.365	-1.572	0.446	0.238	0.377	0.200	0.050	0.050	0.000	0.561
36	0.752	1.433	-1.378	0.392	2.330	1.429	0.890	1.117	0.200	0.040	0.040	0.000	1.092
37	0.413	0.210	-0.097	0.187	-0.693	0.611	0.087	0.605	0.200	0.010	-0.010	0.000	0.456
38	0.823	0.373	-0.197	0.316	-1.003	0.429	0.058	0.425	0.200	0.010	0.010	0.000	0.505
39	0.585	0.442	-0.375	0.233	1.742	1.079	0.142	1.069	0.200	0.030	-0.030	0.000	0.585
40	0.914	0.537	-0.436	0.312	-0.474	0.324	-0.194	0.259	0.200	0.080	-0.080	0.000	0.431
41	0.706	0.338	-0.144	0.306	-1.852	0.867	0.098	0.861	0.200	0.020	0.020	0.000	0.522
42	1.508	0.762	-0.532	0.546	-1.290	0.440	-0.050	0.437	0.200	0.030	-0.030	0.000	0.527
43	0.627	0.570	-0.523	0.227	0.240	0.576	-0.380	0.432	0.200	0.140	-0.140	0.000	0.508
44	0.486	0.275	-0.115	0.250	-1.979	1.293	0.051	1.292	0.200	0.000	0.000	0.000	0.554
45	1.840	1.063	-0.960	0.456	-1.543	0.956	-0.113	0.949	0.200	0.090	-0.090	0.000	0.489
46	1.199	0.608	-0.501	0.345	0.054	0.247	0.134	0.208	0.200	0.120	0.120	0.000	0.782
47	0.653	0.289	-0.147	0.250	2.166	1.175	0.326	1.129	0.200	0.080	0.080	0.000	0.747
48	0.672	0.336	-0.229	0.246	-0.368	0.448	0.202	0.400	0.200	0.090	0.090	0.000	0.597
49	0.711	0.246	-0.079	0.232	-0.092	0.394	0.238	0.314	0.200	0.070	0.070	0.000	0.428
50	0.340	0.152	0.030	0.149	1.330	1.377	-0.110	1.373	0.200	0.010	-0.010	0.000	0.403

Table E.24. Item-Level Evaluation Criteria Summary Statistics, Condition 5, Sample Size = 100

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	0.993	0.417	-0.247	0.336	-2.141	0.915	-0.391	0.827	0.197	0.066	0.067	0.014	0.552
2	0.779	0.475	-0.311	0.359	-0.835	0.857	-0.545	0.662	0.204	0.130	-0.127	0.031	0.519
3	1.098	0.493	-0.272	0.412	-2.438	1.077	-0.558	0.922	0.199	0.058	0.059	0.013	0.596
4	1.055	0.507	-0.215	0.459	-2.512	1.295	-0.562	1.167	0.200	0.048	0.050	0.012	0.624
5	0.650	0.300	-0.141	0.265	-3.920	1.948	-0.980	1.684	0.202	0.020	0.022	0.009	0.414
6	0.872	0.566	-0.389	0.412	-3.793	2.761	-1.263	2.455	0.201	0.019	0.021	0.008	0.695
7	1.411	0.567	-0.259	0.504	-1.570	0.557	-0.190	0.524	0.195	0.055	0.055	0.006	0.539
8	1.991	1.017	-0.729	0.710	-1.374	0.415	-0.124	0.396	0.191	0.061	0.061	0.007	0.686
9	1.861	1.882	-1.789	0.584	-2.059	0.773	-0.609	0.476	0.200	0.113	-0.110	0.027	0.674
10	1.079	0.690	-0.561	0.402	-3.235	1.826	-1.025	1.512	0.201	0.029	0.031	0.010	0.634
11	1.519	0.590	-0.331	0.488	-1.642	0.608	-0.232	0.562	0.195	0.026	0.025	0.008	0.603
12	1.165	0.707	-0.545	0.451	-3.342	2.026	-1.212	1.623	0.202	0.091	-0.088	0.022	0.611
13	1.145	0.544	-0.365	0.403	-2.680	1.415	-0.700	1.230	0.201	0.010	0.011	0.003	0.581
14	1.252	0.712	-0.528	0.478	-1.537	0.795	-0.537	0.587	0.202	0.112	-0.108	0.029	0.516
15	0.645	0.319	-0.125	0.294	-2.709	1.369	-0.589	1.236	0.200	0.049	0.050	0.013	0.473
16	1.012	0.340	-0.168	0.296	-0.928	0.436	0.002	0.436	0.193	0.043	0.043	0.004	0.444
17	1.114	0.975	-0.866	0.447	-3.547	2.879	-1.107	2.658	0.202	0.020	-0.018	0.010	0.740
18	1.416	0.580	-0.384	0.435	-1.854	0.670	-0.314	0.591	0.198	0.028	0.028	0.003	0.554
19	0.919	0.536	-0.351	0.405	-1.146	0.906	-0.566	0.707	0.202	0.112	-0.109	0.027	0.550
20	0.311	0.137	0.051	0.127	-3.250	2.674	0.570	2.613	0.202	0.002	0.002	0.002	0.506
21	1.230	0.484	-0.050	0.481	-0.199	0.380	0.241	0.293	0.188	0.069	0.068	0.008	0.544
22	1.374	0.518	-0.316	0.410	-1.280	0.485	-0.050	0.483	0.191	0.071	0.071	0.007	0.560
23	1.043	0.571	-0.347	0.453	0.399	0.420	0.139	0.397	0.196	0.052	-0.044	0.027	0.565
24	0.833	0.376	-0.137	0.350	-2.311	1.312	-0.441	1.236	0.199	0.038	0.039	0.011	0.554
25	1.525	0.871	-0.695	0.524	-2.225	0.973	-0.555	0.800	0.200	0.011	0.010	0.005	0.772
26	0.854	0.378	-0.136	0.353	-1.105	0.612	-0.225	0.569	0.200	0.025	-0.020	0.015	0.490
27	1.394	0.578	-0.266	0.513	-1.157	0.441	-0.077	0.435	0.196	0.038	0.036	0.013	0.561
28	0.864	0.388	-0.037	0.386	1.057	0.843	0.597	0.596	0.183	0.045	0.043	0.010	0.606
29	1.628	1.057	-0.932	0.499	-2.562	1.144	-0.602	0.973	0.197	0.056	0.057	0.012	0.792
30	1.346	0.513	-0.194	0.475	-0.970	0.351	0.101	0.336	0.190	0.099	0.100	0.010	0.557
31	1.692	0.558	-0.238	0.505	-0.554	0.321	0.166	0.275	0.174	0.114	0.114	0.005	0.607

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.503	0.725	-0.467	0.554	-0.282	0.285	-0.052	0.280	0.200	0.077	-0.070	0.031	0.444
33	1.475	0.469	-0.205	0.422	-0.295	0.353	0.245	0.255	0.175	0.105	0.105	0.005	0.567
34	1.734	1.016	-0.876	0.515	-2.199	0.792	-0.449	0.652	0.197	0.026	0.027	0.007	0.735
35	1.637	0.969	-0.823	0.512	-2.331	0.851	-0.521	0.673	0.197	0.046	0.047	0.011	0.776
36	1.200	1.082	-0.930	0.554	2.588	1.639	1.148	1.169	0.157	0.024	-0.003	0.024	1.059
37	0.457	0.206	-0.053	0.200	-0.908	0.802	-0.128	0.792	0.202	0.013	-0.008	0.010	0.449
38	0.853	0.337	-0.167	0.293	-1.255	0.732	-0.195	0.705	0.199	0.013	0.009	0.010	0.543
39	0.662	0.448	-0.298	0.334	2.656	1.867	1.056	1.539	0.188	0.047	-0.042	0.022	0.682
40	0.974	0.481	-0.376	0.300	-0.453	0.441	-0.173	0.405	0.197	0.087	-0.083	0.024	0.457
41	0.698	0.286	-0.152	0.242	-2.504	1.194	-0.554	1.058	0.201	0.020	0.021	0.008	0.453
42	1.451	0.808	-0.589	0.553	-1.651	0.709	-0.411	0.577	0.199	0.037	-0.031	0.020	0.675
43	0.751	0.547	-0.399	0.375	0.471	0.575	-0.149	0.555	0.200	0.143	-0.140	0.032	0.641
44	0.519	0.249	-0.081	0.236	-2.803	1.723	-0.773	1.539	0.202	0.005	0.002	0.004	0.465
45	1.760	1.190	-1.040	0.578	-2.082	1.084	-0.652	0.865	0.199	0.094	-0.091	0.024	0.677
46	1.644	0.426	-0.057	0.423	0.271	0.423	0.351	0.236	0.165	0.086	0.085	0.009	0.577
47	0.707	0.315	-0.094	0.301	3.359	2.321	1.519	1.755	0.170	0.053	0.050	0.014	0.665
48	0.788	0.322	-0.112	0.302	-0.318	0.607	0.252	0.552	0.194	0.083	0.084	0.012	0.590
49	0.765	0.297	-0.025	0.296	0.042	0.589	0.372	0.457	0.192	0.062	0.062	0.007	0.487
50	0.356	0.153	0.046	0.146	1.695	1.262	0.255	1.236	0.204	0.012	-0.006	0.011	0.409

Table E.25. Item-Level Evaluation Criteria Summary Statistics, Condition 0, Sample Size = 75

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.383	0.607	0.143	0.590	-1.818	0.802	-0.068	0.799	0.198	0.066	0.068	0.014	0.543
2	1.060	0.622	-0.030	0.621	-0.713	0.695	-0.423	0.551	0.205	0.129	-0.125	0.030	0.538
3	1.578	0.740	0.208	0.711	-2.033	1.015	-0.153	1.003	0.200	0.058	0.060	0.014	0.632
4	1.507	0.839	0.237	0.805	-2.093	1.150	-0.143	1.141	0.201	0.050	0.051	0.010	0.704
5	0.885	0.454	0.095	0.444	-3.171	1.596	-0.231	1.580	0.202	0.020	0.022	0.009	0.525
6	1.306	0.752	0.046	0.751	-2.931	2.867	-0.401	2.839	0.202	0.020	0.022	0.009	0.963
7	1.948	0.937	0.278	0.895	-1.330	0.548	0.050	0.546	0.199	0.059	0.059	0.005	0.526
8	3.004	1.291	0.284	1.260	-1.189	0.669	0.061	0.666	0.195	0.065	0.065	0.008	0.573
9	2.911	1.196	-0.739	0.940	-1.696	0.615	-0.246	0.564	0.203	0.110	-0.107	0.028	0.551
10	1.612	0.843	-0.028	0.842	-2.556	2.112	-0.346	2.084	0.202	0.031	0.032	0.009	0.751
11	2.283	0.963	0.433	0.860	-1.302	0.531	0.108	0.520	0.198	0.029	0.028	0.008	0.606
12	1.810	0.819	0.100	0.813	-2.269	2.392	-0.139	2.388	0.204	0.090	-0.086	0.025	1.016
13	1.657	0.778	0.147	0.764	-2.385	1.557	-0.405	1.503	0.202	0.012	0.012	0.005	0.703
14	1.852	0.969	0.072	0.966	-1.264	0.596	-0.264	0.535	0.206	0.109	-0.104	0.033	0.560
15	0.833	0.468	0.063	0.463	-2.351	1.328	-0.231	1.308	0.201	0.050	0.051	0.013	0.523
16	1.328	0.526	0.148	0.505	-0.814	0.428	0.116	0.412	0.197	0.047	0.047	0.003	0.442
17	1.917	0.821	-0.063	0.819	-2.740	2.572	-0.300	2.554	0.203	0.020	-0.017	0.012	0.747
18	2.144	0.956	0.344	0.892	-1.496	0.535	0.044	0.534	0.203	0.036	0.033	0.014	0.585
19	1.174	0.591	-0.096	0.583	-1.013	0.831	-0.433	0.709	0.206	0.108	-0.104	0.028	0.546
20	0.415	0.267	0.155	0.217	-3.052	1.934	0.768	1.775	0.204	0.006	0.004	0.004	0.510
21	1.497	0.700	0.217	0.666	-0.241	0.398	0.199	0.345	0.194	0.074	0.074	0.006	0.532
22	1.877	0.768	0.187	0.745	-1.140	0.605	0.090	0.598	0.196	0.075	0.076	0.007	0.543
23	1.304	0.637	-0.086	0.631	0.325	0.393	0.065	0.388	0.201	0.048	-0.039	0.027	0.490
24	1.130	0.588	0.160	0.566	-1.879	1.044	-0.009	1.044	0.202	0.041	0.042	0.010	0.607
25	2.349	1.091	0.129	1.083	-2.073	1.508	-0.403	1.453	0.206	0.021	0.016	0.015	0.779
26	1.126	0.700	0.136	0.686	-0.984	0.625	-0.104	0.616	0.203	0.022	-0.017	0.015	0.515
27	1.900	1.002	0.240	0.972	-1.054	0.506	0.026	0.505	0.200	0.041	0.040	0.012	0.572
28	1.033	0.541	0.133	0.524	0.888	0.716	0.428	0.574	0.189	0.050	0.049	0.010	0.588
29	2.612	0.732	0.052	0.731	-2.061	1.069	-0.101	1.065	0.199	0.058	0.059	0.012	0.612
30	1.761	0.793	0.221	0.761	-0.885	0.408	0.186	0.364	0.194	0.104	0.104	0.011	0.497
31	2.252	0.958	0.322	0.902	-0.505	0.374	0.215	0.306	0.182	0.121	0.122	0.005	0.566

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	2.038	0.971	0.068	0.969	-0.268	0.294	-0.038	0.292	0.202	0.075	-0.068	0.031	0.466
33	1.851	0.811	0.171	0.793	-0.286	0.359	0.254	0.253	0.184	0.114	0.114	0.001	0.530
34	2.686	0.892	0.076	0.889	-1.818	0.650	-0.068	0.647	0.200	0.030	0.030	0.006	0.585
35	2.552	0.866	0.092	0.861	-1.846	0.628	-0.036	0.627	0.199	0.048	0.049	0.009	0.641
36	1.695	0.981	-0.436	0.879	2.043	1.013	0.603	0.815	0.167	0.025	0.007	0.024	0.782
37	0.603	0.474	0.093	0.465	-0.808	0.871	-0.028	0.870	0.204	0.015	-0.006	0.014	0.536
38	1.129	0.552	0.109	0.542	-1.073	0.628	-0.013	0.628	0.203	0.017	0.013	0.012	0.548
39	0.909	0.471	-0.051	0.469	2.050	1.136	0.450	1.044	0.196	0.041	-0.035	0.021	0.597
40	1.293	0.540	-0.057	0.537	-0.427	0.436	-0.147	0.410	0.203	0.081	-0.077	0.027	0.504
41	0.943	0.425	0.093	0.415	-2.101	1.095	-0.151	1.084	0.202	0.020	0.022	0.008	0.515
42	2.101	0.995	0.061	0.993	-1.435	0.692	-0.195	0.664	0.203	0.035	-0.027	0.023	0.670
43	1.000	0.667	-0.150	0.650	0.367	0.547	-0.253	0.485	0.203	0.141	-0.137	0.033	0.608
44	0.650	0.357	0.050	0.354	-2.423	1.469	-0.393	1.415	0.202	0.009	0.002	0.008	0.538
45	2.704	0.953	-0.096	0.949	-1.724	0.872	-0.294	0.821	0.203	0.091	-0.087	0.026	0.612
46	2.015	0.840	0.315	0.779	0.189	0.370	0.269	0.254	0.176	0.096	0.096	0.006	0.584
47	0.941	0.388	0.141	0.362	2.458	1.786	0.618	1.676	0.182	0.064	0.062	0.015	0.878
48	0.944	0.432	0.044	0.430	-0.324	0.576	0.247	0.520	0.196	0.085	0.086	0.013	0.580
49	0.945	0.479	0.155	0.454	0.013	0.562	0.343	0.445	0.195	0.064	0.065	0.010	0.493
50	0.488	0.319	0.178	0.265	1.458	1.257	0.018	1.257	0.206	0.012	-0.004	0.011	0.573

Table E.26. Item-Level Evaluation Criteria Summary Statistics, Condition 1, Sample Size = 75

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.240	0.585	0.000	0.585	-2.243	1.471	-0.493	1.386	0.199	0.068	0.069	0.015	0.654
2	0.728	0.513	-0.362	0.363	-1.077	1.244	-0.787	0.964	0.198	0.135	-0.132	0.029	0.559
3	1.346	0.597	-0.024	0.596	-2.143	2.845	-0.263	2.833	0.200	0.059	0.060	0.014	0.679
4	1.252	0.600	-0.018	0.600	-2.333	2.629	-0.383	2.601	0.200	0.049	0.050	0.012	0.830
5	0.747	0.478	-0.043	0.476	-2.384	5.345	0.557	5.316	0.202	0.020	0.022	0.009	1.195
6	1.104	0.696	-0.156	0.678	-0.832	8.546	1.698	8.376	0.202	0.020	0.022	0.009	1.497
7	1.495	0.524	-0.175	0.494	-1.628	0.813	-0.248	0.774	0.196	0.056	0.056	0.009	0.600
8	1.822	1.011	-0.898	0.464	-1.214	2.356	0.036	2.356	0.188	0.058	0.058	0.005	0.690
9	1.916	1.790	-1.734	0.447	-2.098	0.901	-0.648	0.627	0.199	0.115	-0.112	0.027	0.672
10	1.302	0.728	-0.338	0.645	-1.059	7.452	1.151	7.362	0.201	0.030	0.031	0.010	1.113
11	1.676	0.490	-0.174	0.459	-1.596	0.687	-0.186	0.661	0.195	0.025	0.025	0.006	0.615
12	1.474	0.677	-0.237	0.634	-1.300	6.484	0.830	6.431	0.203	0.090	-0.087	0.023	1.197
13	1.411	0.653	-0.099	0.646	-1.776	5.208	0.204	5.204	0.202	0.013	0.012	0.006	0.846
14	1.347	0.712	-0.433	0.564	-1.628	0.973	-0.628	0.744	0.200	0.113	-0.110	0.029	0.606
15	0.700	0.421	-0.070	0.415	-1.926	3.883	0.194	3.878	0.201	0.050	0.051	0.014	0.864
16	1.059	0.454	-0.121	0.438	-1.020	0.533	-0.090	0.525	0.193	0.042	0.043	0.009	0.505
17	1.540	0.808	-0.440	0.677	-0.818	7.706	1.622	7.534	0.202	0.021	-0.018	0.011	1.313
18	1.617	0.591	-0.183	0.562	-1.895	0.942	-0.355	0.872	0.199	0.030	0.029	0.005	0.624
19	0.828	0.602	-0.442	0.410	-1.412	1.318	-0.832	1.022	0.199	0.115	-0.111	0.027	0.612
20	0.290	0.169	0.030	0.167	-2.435	5.658	1.385	5.486	0.202	0.002	0.002	0.002	0.914
21	0.993	0.499	-0.288	0.407	-0.397	0.467	0.043	0.465	0.185	0.065	0.065	0.007	0.695
22	1.485	0.576	-0.206	0.538	-1.429	0.976	-0.199	0.955	0.194	0.073	0.074	0.010	0.603
23	0.815	0.661	-0.575	0.325	0.349	0.564	0.089	0.557	0.188	0.056	-0.052	0.021	0.668
24	0.978	0.523	0.008	0.523	-2.021	2.601	-0.151	2.597	0.201	0.039	0.041	0.012	0.680
25	1.646	0.822	-0.574	0.588	-1.461	5.138	0.209	5.134	0.200	0.012	0.010	0.006	1.144
26	0.803	0.432	-0.187	0.389	-1.382	1.178	-0.502	1.065	0.198	0.026	-0.022	0.013	0.593
27	1.338	0.650	-0.322	0.565	-1.404	0.848	-0.324	0.783	0.194	0.034	0.034	0.003	0.687
28	0.678	0.344	-0.222	0.263	1.111	1.035	0.651	0.805	0.182	0.043	0.042	0.009	0.673
29	1.928	0.732	-0.632	0.369	-1.964	3.689	-0.004	3.689	0.198	0.056	0.058	0.012	0.797
30	1.404	0.504	-0.136	0.485	-1.084	0.491	-0.014	0.491	0.192	0.101	0.102	0.014	0.613
31	1.567	0.591	-0.363	0.467	-0.661	0.413	0.059	0.409	0.177	0.116	0.117	0.011	0.706

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.303	0.799	-0.667	0.440	-0.420	0.426	-0.190	0.381	0.189	0.084	-0.081	0.024	0.531
33	1.300	0.594	-0.380	0.457	-0.402	0.341	0.139	0.312	0.179	0.108	0.109	0.010	0.707
34	1.856	0.867	-0.754	0.428	-2.241	1.048	-0.491	0.926	0.197	0.026	0.027	0.006	0.709
35	1.843	0.749	-0.617	0.425	-2.236	0.916	-0.426	0.811	0.197	0.046	0.047	0.011	0.673
36	0.836	1.398	-1.294	0.528	4.019	4.436	2.579	3.609	0.163	0.020	0.003	0.020	1.327
37	0.418	0.258	-0.092	0.241	-1.085	1.928	-0.305	1.903	0.200	0.014	-0.010	0.009	0.579
38	0.875	0.447	-0.145	0.423	-1.381	0.869	-0.321	0.808	0.198	0.009	0.008	0.005	0.624
39	0.594	0.451	-0.366	0.263	2.895	2.403	1.295	2.024	0.188	0.047	-0.043	0.020	0.684
40	0.943	0.552	-0.407	0.373	-0.601	0.603	-0.321	0.511	0.194	0.089	-0.086	0.024	0.564
41	0.777	0.424	-0.073	0.418	-2.204	3.060	-0.254	3.050	0.201	0.019	0.021	0.008	0.682
42	1.553	0.755	-0.487	0.576	-1.846	1.211	-0.606	1.049	0.199	0.035	-0.032	0.016	0.791
43	0.660	0.575	-0.490	0.300	0.425	0.728	-0.195	0.702	0.194	0.150	-0.146	0.032	0.624
44	0.483	0.309	-0.117	0.286	-2.465	4.230	-0.435	4.208	0.201	0.004	0.001	0.004	0.615
45	1.806	1.108	-0.994	0.490	-1.841	2.256	-0.411	2.219	0.198	0.095	-0.092	0.024	0.723
46	1.229	0.627	-0.471	0.413	0.204	0.436	0.284	0.331	0.167	0.086	0.087	0.004	0.794
47	0.681	0.305	-0.119	0.281	3.279	2.215	1.439	1.685	0.178	0.060	0.058	0.014	0.940
48	0.689	0.385	-0.211	0.322	-0.508	0.808	0.062	0.806	0.191	0.080	0.081	0.013	0.704
49	0.732	0.305	-0.058	0.300	-0.034	0.640	0.296	0.567	0.191	0.060	0.061	0.010	0.542
50	0.346	0.206	0.036	0.203	2.193	2.366	0.753	2.243	0.200	0.016	-0.010	0.012	0.510

Table E.27. Item-Level Evaluation Criteria Summary Statistics, Condition 2, Sample Size = 75

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.000	0.240	-0.240	0.000	-1.954	0.495	-0.204	0.451	0.200	0.070	0.070	0.000	0.519
2	1.000	0.090	-0.090	0.000	-0.541	0.431	-0.251	0.350	0.200	0.130	-0.130	0.000	0.360
3	1.000	0.370	-0.370	0.000	-2.314	0.634	-0.434	0.461	0.200	0.060	0.060	0.000	0.578
4	1.000	0.270	-0.270	0.000	-2.213	0.553	-0.263	0.486	0.200	0.050	0.050	0.000	0.499
5	1.000	0.210	0.210	0.000	-2.195	1.722	0.745	1.553	0.200	0.020	0.020	0.000	0.645
6	1.000	0.260	-0.260	0.000	-3.036	0.765	-0.506	0.574	0.200	0.020	0.020	0.000	0.432
7	1.000	0.670	-0.670	0.000	-1.834	0.645	-0.454	0.457	0.200	0.060	0.060	0.000	0.740
8	1.000	1.720	-1.720	0.000	-1.939	0.781	-0.689	0.367	0.200	0.070	0.070	0.000	1.146
9	1.000	2.650	-2.650	0.000	-2.840	1.449	-1.390	0.412	0.200	0.110	-0.110	0.000	1.098
10	1.000	0.640	-0.640	0.000	-2.969	0.925	-0.759	0.528	0.200	0.030	0.030	0.000	0.598
11	1.000	0.850	-0.850	0.000	-1.949	0.667	-0.539	0.394	0.200	0.030	0.030	0.000	0.763
12	1.000	0.710	-0.710	0.000	-3.272	1.356	-1.142	0.731	0.200	0.090	-0.090	0.000	0.635
13	1.000	0.510	-0.510	0.000	-2.611	0.788	-0.631	0.472	0.200	0.010	0.010	0.000	0.571
14	1.000	0.780	-0.780	0.000	-1.630	0.723	-0.630	0.353	0.200	0.110	-0.110	0.000	0.480
15	1.000	0.230	0.230	0.000	-1.604	0.653	0.516	0.400	0.200	0.050	0.050	0.000	0.418
16	1.000	0.180	-0.180	0.000	-0.863	0.345	0.067	0.338	0.200	0.050	0.050	0.000	0.360
17	1.000	0.980	-0.980	0.000	-3.846	1.548	-1.406	0.648	0.200	0.020	-0.020	0.000	0.678
18	1.000	0.800	-0.800	0.000	-2.163	0.783	-0.623	0.475	0.200	0.030	0.030	0.000	0.729
19	1.000	0.270	-0.270	0.000	-0.908	0.551	-0.328	0.443	0.200	0.110	-0.110	0.000	0.354
20	1.000	0.740	0.740	0.000	-1.008	2.867	2.812	0.556	0.200	0.000	0.000	0.000	1.153
21	1.000	0.280	-0.280	0.000	-0.235	0.407	0.205	0.352	0.200	0.080	0.080	0.000	0.515
22	1.000	0.690	-0.690	0.000	-1.490	0.488	-0.260	0.413	0.200	0.080	0.080	0.000	0.724
23	1.000	0.390	-0.390	0.000	0.381	0.393	0.121	0.374	0.200	0.040	-0.040	0.000	0.420
24	1.000	0.030	0.030	0.000	-1.652	0.500	0.218	0.450	0.200	0.040	0.040	0.000	0.346
25	1.000	1.220	-1.220	0.000	-2.761	1.242	-1.091	0.593	0.200	0.010	0.010	0.000	1.081
26	1.000	0.010	0.010	0.000	-0.843	0.328	0.037	0.326	0.200	0.020	-0.020	0.000	0.285
27	1.000	0.660	-0.660	0.000	-1.357	0.481	-0.277	0.394	0.200	0.040	0.040	0.000	0.673
28	1.000	0.100	0.100	0.000	0.914	0.595	0.454	0.386	0.200	0.060	0.060	0.000	0.410
29	1.000	1.560	-1.560	0.000	-3.187	1.353	-1.227	0.572	0.200	0.060	0.060	0.000	1.138
30	1.000	0.540	-0.540	0.000	-1.165	0.384	-0.095	0.372	0.200	0.110	0.110	0.000	0.694
31	1.000	0.930	-0.930	0.000	-0.744	0.365	-0.024	0.365	0.200	0.140	0.140	0.000	0.945

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.000	0.970	-0.970	0.000	-0.422	0.359	-0.192	0.303	0.200	0.070	-0.070	0.000	0.484
33	1.000	0.680	-0.680	0.000	-0.387	0.378	0.153	0.345	0.200	0.130	0.130	0.000	0.804
34	1.000	1.610	-1.610	0.000	-2.900	1.251	-1.150	0.492	0.200	0.030	0.030	0.000	1.140
35	1.000	1.460	-1.460	0.000	-2.901	1.202	-1.091	0.504	0.200	0.050	0.050	0.000	1.066
36	1.000	1.130	-1.130	0.000	2.800	1.620	1.360	0.880	0.200	0.040	0.040	0.000	1.188
37	1.000	0.490	0.490	0.000	-0.235	0.690	0.546	0.423	0.200	0.010	-0.010	0.000	0.736
38	1.000	0.020	-0.020	0.000	-0.960	0.407	0.100	0.394	0.200	0.010	0.010	0.000	0.345
39	1.000	0.040	0.040	0.000	1.751	0.506	0.151	0.483	0.200	0.030	-0.030	0.000	0.467
40	1.000	0.350	-0.350	0.000	-0.417	0.347	-0.137	0.319	0.200	0.080	-0.080	0.000	0.343
41	1.000	0.150	0.150	0.000	-1.662	0.497	0.288	0.405	0.200	0.020	0.020	0.000	0.371
42	1.000	1.040	-1.040	0.000	-1.985	0.871	-0.745	0.451	0.200	0.030	-0.030	0.000	0.844
43	1.000	0.150	-0.150	0.000	0.404	0.430	-0.216	0.371	0.200	0.140	-0.140	0.000	0.510
44	1.000	0.400	0.400	0.000	-1.284	0.897	0.746	0.499	0.200	0.000	0.000	0.000	0.652
45	1.000	1.800	-1.800	0.000	-2.695	1.353	-1.265	0.478	0.200	0.090	-0.090	0.000	1.005
46	1.000	0.700	-0.700	0.000	0.308	0.544	0.388	0.382	0.200	0.120	0.120	0.000	0.855
47	1.000	0.200	0.200	0.000	2.486	0.857	0.646	0.564	0.200	0.080	0.080	0.000	0.514
48	1.000	0.100	0.100	0.000	-0.136	0.574	0.434	0.375	0.200	0.090	0.090	0.000	0.350
49	1.000	0.210	0.210	0.000	0.116	0.613	0.446	0.420	0.200	0.070	0.070	0.000	0.356
50	1.000	0.690	0.690	0.000	0.887	0.708	-0.553	0.441	0.200	0.010	-0.010	0.000	0.973

Table E.28. Item-Level Evaluation Criteria Summary Statistics, Condition 3, Sample Size = 75

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.240	0.585	0.000	0.585	-2.244	1.471	-0.494	1.386	0.200	0.070	0.070	0.000	0.651
2	0.728	0.513	-0.362	0.363	-1.070	1.245	-0.780	0.970	0.200	0.130	-0.130	0.000	0.553
3	1.346	0.597	-0.024	0.596	-2.145	2.849	-0.265	2.836	0.200	0.060	0.060	0.000	0.675
4	1.252	0.600	-0.018	0.600	-2.334	2.631	-0.384	2.602	0.200	0.050	0.050	0.000	0.827
5	0.747	0.478	-0.043	0.476	-2.384	5.349	0.556	5.320	0.200	0.020	0.020	0.000	1.194
6	1.104	0.696	-0.156	0.678	-0.831	8.557	1.699	8.387	0.200	0.020	0.020	0.000	1.496
7	1.495	0.524	-0.175	0.494	-1.629	0.816	-0.249	0.777	0.200	0.060	0.060	0.000	0.599
8	1.822	1.011	-0.898	0.464	-1.208	2.357	0.042	2.357	0.200	0.070	0.070	0.000	0.707
9	1.916	1.790	-1.734	0.447	-2.101	0.904	-0.651	0.628	0.200	0.110	-0.110	0.000	0.676
10	1.302	0.728	-0.338	0.645	-1.063	7.458	1.147	7.369	0.200	0.030	0.030	0.000	1.113
11	1.676	0.490	-0.174	0.459	-1.596	0.692	-0.186	0.667	0.200	0.030	0.030	0.000	0.613
12	1.474	0.677	-0.237	0.634	-1.306	6.489	0.824	6.436	0.200	0.090	-0.090	0.000	1.197
13	1.411	0.653	-0.099	0.646	-1.782	5.214	0.198	5.211	0.200	0.010	0.010	0.000	0.843
14	1.347	0.712	-0.433	0.564	-1.630	0.978	-0.630	0.748	0.200	0.110	-0.110	0.000	0.602
15	0.700	0.421	-0.070	0.415	-1.924	3.894	0.196	3.889	0.200	0.050	0.050	0.000	0.863
16	1.059	0.454	-0.121	0.438	-1.009	0.538	-0.079	0.532	0.200	0.050	0.050	0.000	0.502
17	1.540	0.808	-0.440	0.677	-0.823	7.712	1.617	7.541	0.200	0.020	-0.020	0.000	1.313
18	1.617	0.591	-0.183	0.562	-1.900	0.945	-0.360	0.874	0.200	0.030	0.030	0.000	0.622
19	0.828	0.602	-0.442	0.410	-1.410	1.321	-0.830	1.028	0.200	0.110	-0.110	0.000	0.609
20	0.290	0.169	0.030	0.167	-2.442	5.672	1.378	5.502	0.200	0.000	0.000	0.000	0.914
21	0.993	0.499	-0.288	0.407	-0.364	0.486	0.077	0.480	0.200	0.080	0.080	0.000	0.705
22	1.485	0.576	-0.206	0.538	-1.424	0.978	-0.194	0.958	0.200	0.080	0.080	0.000	0.603
23	0.815	0.661	-0.575	0.325	0.390	0.577	0.130	0.562	0.200	0.040	-0.040	0.000	0.667
24	0.978	0.523	0.008	0.523	-2.024	2.606	-0.154	2.601	0.200	0.040	0.040	0.000	0.678
25	1.646	0.822	-0.574	0.588	-1.465	5.143	0.205	5.139	0.200	0.010	0.010	0.000	1.145
26	0.803	0.432	-0.187	0.389	-1.377	1.190	-0.497	1.081	0.200	0.020	-0.020	0.000	0.591
27	1.338	0.650	-0.322	0.565	-1.398	0.853	-0.318	0.792	0.200	0.040	0.040	0.000	0.688
28	0.678	0.344	-0.222	0.263	1.195	1.088	0.735	0.802	0.200	0.060	0.060	0.000	0.684
29	1.928	0.732	-0.632	0.369	-1.966	3.693	-0.006	3.693	0.200	0.060	0.060	0.000	0.801
30	1.404	0.504	-0.136	0.485	-1.073	0.496	-0.003	0.496	0.200	0.110	0.110	0.000	0.619
31	1.567	0.591	-0.363	0.467	-0.632	0.429	0.088	0.420	0.200	0.140	0.140	0.000	0.745

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.303	0.799	-0.667	0.440	-0.403	0.427	-0.173	0.390	0.200	0.070	-0.070	0.000	0.523
33	1.300	0.594	-0.380	0.457	-0.366	0.369	0.174	0.326	0.200	0.130	0.130	0.000	0.740
34	1.856	0.867	-0.754	0.428	-2.244	1.050	-0.494	0.927	0.200	0.030	0.030	0.000	0.715
35	1.843	0.749	-0.617	0.425	-2.238	0.919	-0.428	0.813	0.200	0.050	0.050	0.000	0.676
36	0.836	1.398	-1.294	0.528	4.421	4.899	2.981	3.887	0.200	0.040	0.040	0.000	1.394
37	0.418	0.258	-0.092	0.241	-1.087	1.936	-0.307	1.912	0.200	0.010	-0.010	0.000	0.574
38	0.875	0.447	-0.145	0.423	-1.375	0.873	-0.315	0.814	0.200	0.010	0.010	0.000	0.619
39	0.594	0.451	-0.366	0.263	2.989	2.417	1.389	1.978	0.200	0.030	-0.030	0.000	0.687
40	0.943	0.552	-0.407	0.373	-0.587	0.604	-0.307	0.520	0.200	0.080	-0.080	0.000	0.554
41	0.777	0.424	-0.073	0.418	-2.206	3.064	-0.256	3.054	0.200	0.020	0.020	0.000	0.681
42	1.553	0.755	-0.487	0.576	-1.849	1.217	-0.609	1.053	0.200	0.030	-0.030	0.000	0.791
43	0.660	0.575	-0.490	0.300	0.448	0.723	-0.172	0.702	0.200	0.140	-0.140	0.000	0.605
44	0.483	0.309	-0.117	0.286	-2.465	4.238	-0.435	4.216	0.200	0.000	0.000	0.000	0.611
45	1.806	1.108	-0.994	0.490	-1.844	2.259	-0.414	2.220	0.200	0.090	-0.090	0.000	0.726
46	1.229	0.627	-0.471	0.413	0.276	0.501	0.356	0.352	0.200	0.120	0.120	0.000	0.851
47	0.681	0.305	-0.119	0.281	3.484	2.326	1.644	1.645	0.200	0.080	0.080	0.000	0.967
48	0.689	0.385	-0.211	0.322	-0.482	0.821	0.088	0.816	0.200	0.090	0.090	0.000	0.705
49	0.732	0.305	-0.058	0.300	-0.002	0.664	0.328	0.578	0.200	0.070	0.070	0.000	0.537
50	0.346	0.206	0.036	0.203	2.177	2.324	0.737	2.204	0.200	0.010	-0.010	0.000	0.499

Table E.29. Item-Level Evaluation Criteria Summary Statistics, Condition 4, Sample Size = 75

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.241	0.585	0.001	0.585	-1.582	0.826	0.168	0.808	0.200	0.070	0.070	0.000	0.597
2	0.728	0.512	-0.362	0.363	-0.759	0.734	-0.469	0.565	0.200	0.130	-0.130	0.000	0.485
3	1.346	0.596	-0.024	0.596	-1.685	0.807	0.195	0.784	0.200	0.060	0.060	0.000	0.636
4	1.252	0.600	-0.018	0.600	-1.756	0.939	0.195	0.918	0.200	0.050	0.050	0.000	0.734
5	0.747	0.478	-0.043	0.476	-2.636	1.649	0.304	1.621	0.200	0.020	0.020	0.000	0.761
6	1.104	0.695	-0.156	0.678	-2.724	1.884	-0.194	1.874	0.200	0.020	0.020	0.000	0.824
7	1.495	0.523	-0.175	0.493	-1.213	0.505	0.167	0.476	0.200	0.060	0.060	0.000	0.484
8	1.822	1.011	-0.899	0.464	-1.210	1.174	0.040	1.173	0.200	0.070	0.070	0.000	0.559
9	1.916	1.790	-1.734	0.447	-1.494	0.399	-0.044	0.396	0.200	0.110	-0.110	0.000	0.475
10	1.302	0.728	-0.338	0.645	-2.405	2.186	-0.195	2.177	0.200	0.030	0.030	0.000	0.728
11	1.676	0.491	-0.174	0.459	-1.191	0.470	0.219	0.415	0.200	0.030	0.030	0.000	0.535
12	1.473	0.677	-0.237	0.634	-2.398	2.059	-0.268	2.041	0.200	0.090	-0.090	0.000	0.726
13	1.411	0.653	-0.099	0.646	-1.843	0.984	0.137	0.974	0.200	0.010	0.010	0.000	0.748
14	1.347	0.711	-0.433	0.564	-1.200	0.488	-0.200	0.446	0.200	0.110	-0.110	0.000	0.528
15	0.700	0.421	-0.070	0.415	-1.952	1.345	0.168	1.335	0.200	0.050	0.050	0.000	0.625
16	1.058	0.454	-0.122	0.438	-0.788	0.359	0.142	0.330	0.200	0.050	0.050	0.000	0.452
17	1.540	0.808	-0.440	0.677	-2.894	2.564	-0.454	2.523	0.200	0.020	-0.020	0.000	0.758
18	1.617	0.591	-0.183	0.562	-1.382	0.544	0.158	0.520	0.200	0.030	0.030	0.000	0.588
19	0.828	0.603	-0.442	0.410	-1.000	0.736	-0.420	0.604	0.200	0.110	-0.110	0.000	0.538
20	0.290	0.169	0.030	0.167	-2.148	2.199	1.672	1.429	0.200	0.000	0.000	0.000	0.674
21	0.993	0.499	-0.287	0.407	-0.338	0.319	0.103	0.302	0.200	0.080	0.080	0.000	0.633
22	1.485	0.576	-0.205	0.538	-1.073	0.591	0.157	0.569	0.200	0.080	0.080	0.000	0.536
23	0.815	0.661	-0.575	0.325	0.198	0.429	-0.062	0.424	0.200	0.040	-0.040	0.000	0.588
24	0.979	0.523	0.009	0.522	-1.562	0.910	0.309	0.856	0.200	0.040	0.040	0.000	0.624
25	1.646	0.822	-0.574	0.588	-1.807	1.383	-0.137	1.376	0.200	0.010	0.010	0.000	0.634
26	0.803	0.432	-0.187	0.389	-0.978	0.651	-0.098	0.643	0.200	0.020	-0.020	0.000	0.517
27	1.338	0.650	-0.322	0.565	-1.043	0.477	0.037	0.476	0.200	0.040	0.040	0.000	0.570
28	0.679	0.343	-0.221	0.262	0.707	0.613	0.247	0.561	0.200	0.060	0.060	0.000	0.613
29	1.928	0.732	-0.632	0.369	-1.650	0.628	0.310	0.547	0.200	0.060	0.060	0.000	0.651
30	1.403	0.504	-0.137	0.485	-0.851	0.383	0.219	0.315	0.200	0.110	0.110	0.000	0.528
31	1.567	0.591	-0.363	0.466	-0.556	0.315	0.164	0.268	0.200	0.140	0.140	0.000	0.664

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.304	0.799	-0.666	0.440	-0.390	0.266	-0.160	0.213	0.200	0.070	-0.070	0.000	0.466
33	1.301	0.594	-0.380	0.457	-0.371	0.277	0.169	0.219	0.200	0.130	0.130	0.000	0.700
34	1.856	0.867	-0.754	0.428	-1.583	0.581	0.167	0.557	0.200	0.030	0.030	0.000	0.543
35	1.843	0.749	-0.617	0.425	-1.585	0.560	0.225	0.512	0.200	0.050	0.050	0.000	0.632
36	0.837	1.397	-1.293	0.528	2.228	1.459	0.788	1.228	0.200	0.040	0.040	0.000	1.125
37	0.418	0.258	-0.092	0.241	-0.660	0.748	0.121	0.738	0.200	0.010	-0.010	0.000	0.535
38	0.876	0.447	-0.144	0.423	-0.997	0.494	0.063	0.490	0.200	0.010	0.010	0.000	0.534
39	0.594	0.451	-0.366	0.263	1.792	1.225	0.192	1.210	0.200	0.030	-0.030	0.000	0.586
40	0.943	0.552	-0.407	0.373	-0.486	0.376	-0.206	0.315	0.200	0.080	-0.080	0.000	0.490
41	0.777	0.424	-0.073	0.418	-1.811	0.976	0.139	0.966	0.200	0.020	0.020	0.000	0.586
42	1.553	0.755	-0.487	0.577	-1.336	0.625	-0.096	0.617	0.200	0.030	-0.030	0.000	0.588
43	0.660	0.575	-0.490	0.300	0.228	0.587	-0.392	0.437	0.200	0.140	-0.140	0.000	0.521
44	0.482	0.310	-0.118	0.286	-2.080	1.317	-0.050	1.316	0.200	0.000	0.000	0.000	0.569
45	1.806	1.108	-0.994	0.489	-1.573	1.028	-0.143	1.018	0.200	0.090	-0.090	0.000	0.518
46	1.228	0.627	-0.472	0.413	0.073	0.297	0.153	0.255	0.200	0.120	0.120	0.000	0.816
47	0.681	0.305	-0.119	0.281	2.109	1.311	0.269	1.283	0.200	0.080	0.080	0.000	0.967
48	0.689	0.385	-0.211	0.321	-0.385	0.534	0.185	0.501	0.200	0.090	0.090	0.000	0.632
49	0.732	0.305	-0.058	0.299	-0.076	0.461	0.254	0.385	0.200	0.070	0.070	0.000	0.471
50	0.346	0.206	0.036	0.202	1.165	1.281	-0.275	1.251	0.200	0.010	-0.010	0.000	0.479

Table E.30. Item-Level Evaluation Criteria Summary Statistics, Condition 5, Sample Size = 75

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	0.958	0.459	-0.282	0.362	-2.278	1.158	-0.528	1.031	0.198	0.066	0.068	0.015	0.626
2	0.803	0.519	-0.287	0.432	-0.823	0.847	-0.533	0.658	0.202	0.131	-0.128	0.030	0.524
3	1.048	0.528	-0.322	0.418	-2.571	1.401	-0.691	1.219	0.199	0.058	0.059	0.014	0.661
4	1.006	0.516	-0.264	0.443	-2.635	1.553	-0.685	1.394	0.199	0.048	0.049	0.012	0.693
5	0.633	0.318	-0.157	0.277	-3.958	2.118	-1.018	1.858	0.202	0.020	0.022	0.009	0.431
6	0.820	0.595	-0.440	0.401	-3.299	3.860	-0.769	3.782	0.202	0.020	0.022	0.009	1.098
7	1.299	0.633	-0.371	0.513	-1.669	0.800	-0.289	0.746	0.196	0.055	0.056	0.009	0.623
8	1.881	1.116	-0.839	0.735	-1.466	0.817	-0.216	0.788	0.190	0.060	0.060	0.006	0.724
9	1.661	2.078	-1.989	0.602	-2.246	1.185	-0.796	0.878	0.199	0.114	-0.111	0.026	0.736
10	0.999	0.793	-0.641	0.466	-2.637	3.197	-0.427	3.169	0.202	0.030	0.032	0.011	0.957
11	1.478	0.630	-0.372	0.509	-1.636	0.741	-0.226	0.705	0.195	0.025	0.025	0.004	0.653
12	1.044	0.792	-0.666	0.428	-3.186	2.996	-1.056	2.804	0.202	0.091	-0.088	0.022	1.015
13	1.064	0.635	-0.446	0.451	-2.828	2.308	-0.848	2.146	0.200	0.010	0.010	0.002	0.699
14	1.210	0.740	-0.570	0.472	-1.571	0.875	-0.571	0.663	0.201	0.113	-0.109	0.029	0.567
15	0.621	0.336	-0.149	0.301	-2.855	1.679	-0.735	1.510	0.201	0.049	0.051	0.013	0.509
16	0.985	0.388	-0.195	0.336	-0.950	0.539	-0.020	0.538	0.195	0.044	0.045	0.006	0.499
17	0.994	1.075	-0.986	0.427	-3.660	3.474	-1.220	3.253	0.202	0.020	-0.018	0.010	0.854
18	1.336	0.655	-0.464	0.463	-1.914	0.796	-0.374	0.703	0.198	0.027	0.028	0.006	0.623
19	0.878	0.562	-0.392	0.402	-1.196	1.057	-0.616	0.859	0.202	0.111	-0.108	0.027	0.579
20	0.335	0.174	0.075	0.157	-3.287	2.411	0.533	2.351	0.203	0.003	0.003	0.001	0.538
21	1.142	0.478	-0.138	0.457	-0.223	0.479	0.217	0.427	0.190	0.070	0.070	0.006	0.634
22	1.291	0.616	-0.399	0.469	-1.392	0.788	-0.162	0.771	0.193	0.072	0.073	0.009	0.651
23	0.997	0.610	-0.393	0.467	0.457	0.516	0.197	0.477	0.196	0.051	-0.044	0.026	0.617
24	0.814	0.391	-0.156	0.359	-2.312	1.328	-0.442	1.252	0.200	0.039	0.040	0.011	0.595
25	1.364	1.039	-0.856	0.589	-2.452	2.193	-0.782	2.049	0.200	0.011	0.010	0.005	0.960
26	0.842	0.433	-0.148	0.406	-1.146	0.769	-0.266	0.721	0.201	0.023	-0.019	0.013	0.541
27	1.304	0.654	-0.356	0.548	-1.258	0.638	-0.178	0.612	0.196	0.037	0.036	0.008	0.669
28	0.818	0.416	-0.082	0.408	1.141	0.992	0.681	0.722	0.184	0.046	0.044	0.012	0.682
29	1.465	1.209	-1.096	0.510	-2.575	1.997	-0.615	1.899	0.198	0.057	0.058	0.013	0.945
30	1.247	0.526	-0.293	0.437	-1.044	0.462	0.026	0.461	0.191	0.100	0.101	0.013	0.644
31	1.605	0.648	-0.325	0.561	-0.548	0.422	0.172	0.385	0.179	0.119	0.119	0.006	0.687

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.460	0.801	-0.510	0.618	-0.270	0.341	-0.040	0.338	0.196	0.080	-0.074	0.029	0.512
33	1.375	0.574	-0.305	0.486	-0.276	0.405	0.264	0.307	0.181	0.111	0.111	0.002	0.657
34	1.549	1.194	-1.061	0.549	-2.436	1.264	-0.686	1.061	0.198	0.027	0.028	0.008	0.844
35	1.480	1.115	-0.980	0.531	-2.511	1.330	-0.701	1.130	0.198	0.046	0.048	0.012	0.830
36	1.036	1.228	-1.095	0.556	3.004	2.150	1.564	1.475	0.161	0.023	0.001	0.023	1.169
37	0.459	0.247	-0.052	0.242	-0.900	1.020	-0.120	1.013	0.202	0.013	-0.008	0.011	0.529
38	0.846	0.402	-0.175	0.362	-1.277	0.821	-0.217	0.791	0.200	0.014	0.010	0.009	0.593
39	0.634	0.455	-0.326	0.317	2.700	1.731	1.100	1.336	0.187	0.048	-0.043	0.021	0.704
40	0.964	0.532	-0.386	0.365	-0.471	0.548	-0.191	0.514	0.198	0.085	-0.082	0.025	0.556
41	0.704	0.316	-0.146	0.281	-2.543	1.348	-0.593	1.211	0.201	0.020	0.021	0.008	0.500
42	1.343	0.884	-0.697	0.544	-1.816	1.040	-0.576	0.865	0.199	0.036	-0.031	0.018	0.794
43	0.760	0.601	-0.390	0.457	0.508	0.601	-0.112	0.591	0.197	0.147	-0.143	0.033	0.662
44	0.508	0.269	-0.092	0.253	-2.848	1.782	-0.818	1.583	0.202	0.007	0.002	0.006	0.518
45	1.597	1.335	-1.203	0.579	-2.276	1.448	-0.846	1.175	0.198	0.095	-0.092	0.024	0.762
46	1.522	0.537	-0.178	0.506	0.304	0.501	0.384	0.322	0.170	0.091	0.090	0.010	0.725
47	0.630	0.321	-0.170	0.272	3.678	2.605	1.838	1.845	0.174	0.055	0.054	0.013	0.712
48	0.744	0.356	-0.156	0.320	-0.322	0.666	0.248	0.618	0.194	0.083	0.084	0.012	0.643
49	0.736	0.331	-0.054	0.326	0.081	0.686	0.411	0.550	0.192	0.061	0.062	0.010	0.551
50	0.386	0.211	0.076	0.197	1.798	1.482	0.358	1.438	0.204	0.013	-0.006	0.011	0.486

Table E.31. Item-Level Evaluation Criteria Summary Statistics, Condition 0, Sample Size = 50

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.464	0.690	0.224	0.652	-1.602	0.983	0.148	0.972	0.197	0.067	0.067	0.008	0.526
2	1.075	0.849	-0.015	0.849	-0.713	0.783	-0.423	0.659	0.201	0.130	-0.129	0.012	0.473
3	1.684	0.978	0.314	0.926	-2.020	1.350	-0.140	1.343	0.197	0.058	0.057	0.007	0.562
4	1.737	0.970	0.467	0.851	-1.830	1.049	0.120	1.042	0.199	0.049	0.049	0.006	0.530
5	0.926	0.614	0.136	0.598	-3.286	1.791	-0.346	1.757	0.200	0.020	0.020	0.003	0.437
6	1.361	0.875	0.101	0.869	-2.845	3.303	-0.315	3.288	0.200	0.020	0.020	0.004	0.706
7	2.203	1.329	0.533	1.217	-1.166	0.623	0.214	0.585	0.198	0.060	0.058	0.015	0.475
8	3.043	1.467	0.323	1.432	-1.134	0.741	0.117	0.732	0.194	0.066	0.064	0.015	0.530
9	2.943	1.189	-0.707	0.956	-1.505	0.409	-0.055	0.405	0.200	0.111	-0.110	0.012	0.454
10	1.742	0.964	0.102	0.959	-2.485	1.341	-0.275	1.312	0.201	0.032	0.031	0.007	0.547
11	2.445	1.218	0.595	1.063	-1.215	0.624	0.195	0.593	0.196	0.030	0.026	0.014	0.507
12	1.896	0.937	0.186	0.918	-2.200	2.238	-0.070	2.237	0.201	0.090	-0.090	0.011	0.903
13	1.664	0.820	0.154	0.805	-2.393	1.743	-0.413	1.693	0.198	0.012	0.008	0.008	0.610
14	1.805	1.063	0.025	1.062	-1.376	1.009	-0.376	0.936	0.203	0.108	-0.107	0.016	0.590
15	0.962	0.749	0.192	0.724	-2.233	1.332	-0.113	1.327	0.199	0.049	0.049	0.007	0.519
16	1.313	0.688	0.133	0.675	-0.750	0.530	0.181	0.498	0.195	0.048	0.045	0.014	0.391
17	2.002	0.851	0.022	0.851	-2.684	1.277	-0.244	1.253	0.201	0.021	-0.019	0.009	0.507
18	2.324	1.234	0.524	1.118	-1.421	0.662	0.119	0.651	0.200	0.032	0.030	0.013	0.505
19	1.183	0.575	-0.087	0.568	-0.935	0.873	-0.355	0.797	0.200	0.110	-0.110	0.011	0.501
20	0.473	0.327	0.213	0.249	-2.495	2.138	1.325	1.678	0.201	0.004	0.001	0.004	0.491
21	1.568	1.063	0.288	1.023	-0.174	0.496	0.266	0.419	0.188	0.069	0.068	0.013	0.529
22	1.903	0.909	0.213	0.883	-1.039	0.604	0.191	0.573	0.195	0.076	0.075	0.014	0.481
23	1.315	0.691	-0.075	0.687	0.338	0.464	0.078	0.457	0.194	0.050	-0.046	0.020	0.472
24	1.273	0.791	0.303	0.731	-1.719	1.559	0.151	1.552	0.200	0.041	0.040	0.010	0.639
25	2.591	1.296	0.371	1.242	-2.000	1.669	-0.330	1.636	0.203	0.018	0.013	0.013	0.618
26	1.217	0.849	0.227	0.818	-0.979	0.886	-0.099	0.880	0.202	0.022	-0.018	0.011	0.527
27	1.988	1.400	0.328	1.361	-1.181	1.004	-0.101	0.999	0.198	0.040	0.038	0.015	0.641
28	1.033	0.569	0.133	0.553	0.989	0.821	0.529	0.628	0.189	0.052	0.049	0.015	0.552
29	2.548	0.795	-0.012	0.795	-1.934	1.096	0.027	1.096	0.197	0.058	0.057	0.009	0.566
30	1.853	0.915	0.313	0.859	-0.735	0.466	0.335	0.324	0.193	0.103	0.103	0.013	0.410
31	2.451	1.245	0.521	1.131	-0.388	0.454	0.332	0.310	0.182	0.123	0.122	0.019	0.551

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	2.304	1.575	0.334	1.539	-0.210	0.415	0.020	0.415	0.199	0.074	-0.071	0.019	0.473
33	1.803	0.792	0.123	0.783	-0.246	0.464	0.294	0.359	0.187	0.119	0.117	0.019	0.541
34	2.659	0.880	0.049	0.878	-1.724	0.674	0.026	0.674	0.197	0.028	0.027	0.008	0.509
35	2.391	0.920	-0.069	0.917	-1.896	1.170	-0.086	1.167	0.197	0.048	0.047	0.009	0.652
36	1.487	0.979	-0.643	0.738	2.310	1.375	0.870	1.064	0.172	0.023	0.012	0.020	0.737
37	0.613	0.388	0.103	0.374	-0.623	0.987	0.158	0.974	0.200	0.013	-0.010	0.008	0.457
38	1.265	0.885	0.245	0.850	-1.071	1.015	-0.011	1.015	0.201	0.019	0.011	0.015	0.565
39	0.844	0.565	-0.117	0.553	2.527	1.682	0.927	1.404	0.194	0.040	-0.037	0.015	0.587
40	1.475	0.921	0.125	0.912	-0.374	0.646	-0.094	0.639	0.201	0.081	-0.080	0.017	0.512
41	0.979	0.487	0.129	0.470	-2.135	1.420	-0.185	1.407	0.200	0.021	0.020	0.006	0.521
42	2.154	1.186	0.114	1.180	-1.503	1.358	-0.263	1.333	0.199	0.034	-0.031	0.014	0.590
43	1.059	0.764	-0.091	0.759	0.476	0.465	-0.144	0.442	0.199	0.141	-0.141	0.012	0.474
44	0.725	0.430	0.125	0.412	-2.144	1.291	-0.114	1.286	0.200	0.005	0.000	0.005	0.438
45	2.826	0.949	0.026	0.949	-1.600	0.733	-0.170	0.713	0.199	0.092	-0.091	0.010	0.504
46	2.025	0.964	0.325	0.908	0.271	0.435	0.351	0.257	0.179	0.101	0.099	0.020	0.517
47	0.962	0.496	0.162	0.469	2.680	1.484	0.840	1.223	0.181	0.064	0.061	0.019	0.502
48	0.972	0.491	0.072	0.486	-0.261	0.702	0.310	0.630	0.194	0.085	0.084	0.010	0.484
49	0.995	0.596	0.205	0.559	-0.033	0.571	0.297	0.487	0.192	0.063	0.062	0.013	0.447
50	0.598	0.493	0.288	0.400	1.539	1.290	0.099	1.286	0.201	0.014	-0.009	0.011	0.592

Table E.32. Item-Level Evaluation Criteria Summary Statistics, Condition 1, Sample Size = 50

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.311	0.566	0.071	0.562	-1.504	2.187	0.246	2.173	0.197	0.068	0.067	0.005	0.732
2	0.701	0.540	-0.389	0.374	-1.074	1.349	-0.784	1.098	0.196	0.135	-0.134	0.008	0.483
3	1.365	0.636	-0.005	0.636	-1.290	4.668	0.590	4.631	0.198	0.058	0.058	0.007	0.730
4	1.459	0.658	0.189	0.631	-1.511	2.575	0.439	2.537	0.199	0.049	0.049	0.006	0.793
5	0.793	0.577	0.003	0.577	-0.691	7.506	2.249	7.161	0.200	0.020	0.020	0.002	1.293
6	1.133	0.763	-0.127	0.752	0.347	10.429	2.877	10.024	0.200	0.020	0.020	0.004	1.062
7	1.533	0.522	-0.137	0.503	-1.406	0.722	-0.026	0.722	0.195	0.056	0.055	0.011	0.417
8	1.751	1.067	-0.969	0.446	-1.067	2.232	0.183	2.224	0.188	0.060	0.058	0.011	0.593
9	1.890	1.784	-1.760	0.290	-1.855	0.589	-0.405	0.429	0.197	0.114	-0.114	0.009	0.465
10	1.360	0.732	-0.280	0.677	-1.959	3.855	0.251	3.846	0.200	0.030	0.030	0.006	0.724
11	1.686	0.494	-0.164	0.466	-1.495	0.719	-0.085	0.714	0.193	0.025	0.023	0.010	0.457
12	1.498	0.661	-0.212	0.626	-1.323	5.816	0.808	5.760	0.200	0.091	-0.090	0.006	0.938
13	1.417	0.643	-0.093	0.636	-1.118	6.397	0.862	6.339	0.200	0.012	0.010	0.008	0.828
14	1.284	0.763	-0.496	0.580	-1.821	1.758	-0.821	1.555	0.199	0.112	-0.111	0.011	0.620
15	0.790	0.540	0.020	0.539	-0.889	4.959	1.231	4.804	0.198	0.049	0.048	0.005	0.918
16	1.050	0.497	-0.131	0.480	-0.896	0.612	0.034	0.611	0.192	0.043	0.042	0.010	0.418
17	1.592	0.721	-0.388	0.608	-2.451	3.277	-0.011	3.277	0.201	0.020	-0.020	0.005	0.593
18	1.610	0.562	-0.190	0.528	-1.438	1.909	0.102	1.906	0.196	0.027	0.026	0.009	0.593
19	0.887	0.605	-0.384	0.468	-1.436	1.961	-0.856	1.764	0.196	0.115	-0.114	0.008	0.566
20	0.345	0.230	0.085	0.214	-2.398	4.814	1.422	4.600	0.200	0.003	0.000	0.003	0.596
21	1.092	0.536	-0.188	0.502	-0.255	0.580	0.185	0.550	0.184	0.065	0.064	0.012	0.589
22	1.482	0.603	-0.208	0.566	-1.295	0.917	-0.065	0.914	0.194	0.074	0.074	0.010	0.510
23	0.843	0.680	-0.548	0.404	0.403	0.636	0.143	0.620	0.186	0.056	-0.054	0.014	0.585
24	1.049	0.559	0.079	0.554	-0.617	5.427	1.253	5.281	0.198	0.039	0.038	0.007	0.764
25	1.690	0.769	-0.530	0.557	-0.699	5.987	0.971	5.908	0.198	0.011	0.008	0.008	0.760
26	0.829	0.510	-0.161	0.484	-1.136	2.226	-0.256	2.211	0.196	0.025	-0.024	0.007	0.633
27	1.270	0.742	-0.390	0.632	-0.849	3.074	0.231	3.066	0.193	0.035	0.033	0.011	0.839
28	0.706	0.378	-0.194	0.325	1.265	1.230	0.805	0.930	0.184	0.046	0.044	0.013	0.642
29	1.889	0.763	-0.671	0.363	-1.531	2.752	0.429	2.719	0.196	0.056	0.056	0.007	0.726
30	1.436	0.498	-0.104	0.487	-0.881	0.451	0.189	0.410	0.189	0.100	0.099	0.011	0.430
31	1.589	0.594	-0.341	0.486	-0.501	0.443	0.219	0.386	0.177	0.118	0.117	0.015	0.626

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.306	0.834	-0.664	0.504	-0.354	0.570	-0.124	0.556	0.189	0.082	-0.081	0.013	0.521
33	1.310	0.648	-0.370	0.532	-0.330	0.505	0.210	0.460	0.182	0.113	0.112	0.015	0.674
34	1.850	0.843	-0.760	0.364	-2.093	0.932	-0.343	0.867	0.195	0.026	0.025	0.007	0.510
35	1.730	0.872	-0.730	0.476	-1.645	3.689	0.165	3.685	0.196	0.047	0.046	0.008	0.789
36	0.964	1.301	-1.166	0.576	4.023	4.799	2.583	4.044	0.170	0.020	0.010	0.017	1.006
37	0.457	0.335	-0.053	0.331	-0.970	1.579	-0.190	1.567	0.197	0.015	-0.013	0.007	0.499
38	0.943	0.555	-0.077	0.550	-1.333	1.139	-0.273	1.106	0.197	0.011	0.007	0.008	0.586
39	0.543	0.535	-0.417	0.335	4.256	4.512	2.656	3.647	0.189	0.043	-0.041	0.014	0.656
40	0.981	0.574	-0.369	0.440	-0.571	1.043	-0.291	1.001	0.193	0.087	-0.087	0.009	0.525
41	0.803	0.488	-0.047	0.486	-0.637	5.966	1.313	5.819	0.199	0.020	0.019	0.005	0.921
42	1.501	0.802	-0.539	0.593	-0.692	4.905	0.548	4.874	0.195	0.036	-0.035	0.010	0.793
43	0.677	0.565	-0.473	0.308	0.615	0.709	-0.005	0.709	0.192	0.148	-0.148	0.009	0.482
44	0.562	0.370	-0.038	0.368	-2.804	2.905	-0.774	2.800	0.199	0.004	-0.001	0.004	0.438
45	1.798	1.093	-1.003	0.434	-2.015	1.047	-0.585	0.869	0.196	0.095	-0.094	0.008	0.520
46	1.325	0.609	-0.375	0.479	0.315	0.512	0.395	0.326	0.173	0.094	0.093	0.016	0.654
47	0.776	0.449	-0.024	0.449	3.901	3.904	2.061	3.316	0.180	0.062	0.060	0.018	0.574
48	0.763	0.453	-0.137	0.431	-0.433	1.258	0.137	1.251	0.191	0.082	0.081	0.009	0.588
49	0.776	0.405	-0.014	0.404	-0.081	0.665	0.249	0.617	0.189	0.060	0.059	0.012	0.497
50	0.398	0.276	0.088	0.262	2.494	2.843	1.054	2.641	0.196	0.017	-0.014	0.009	0.499

Table E.33. Item-Level Evaluation Criteria Summary Statistics, Condition 2, Sample Size = 50

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.000	0.240	-0.240	0.000	-1.786	0.574	-0.036	0.573	0.200	0.070	0.070	0.000	0.438
2	1.000	0.090	-0.090	0.000	-0.469	0.450	-0.179	0.413	0.200	0.130	-0.130	0.000	0.340
3	1.000	0.370	-0.370	0.000	-2.221	0.660	-0.341	0.565	0.200	0.060	0.060	0.000	0.474
4	1.000	0.270	-0.270	0.000	-2.143	0.537	-0.193	0.501	0.200	0.050	0.050	0.000	0.384
5	1.000	0.210	0.210	0.000	-2.348	0.872	0.592	0.641	0.200	0.020	0.020	0.000	0.404
6	1.000	0.260	-0.260	0.000	-2.979	0.755	-0.449	0.608	0.200	0.020	0.020	0.000	0.322
7	1.000	0.670	-0.670	0.000	-1.653	0.522	-0.273	0.445	0.200	0.060	0.060	0.000	0.535
8	1.000	1.720	-1.720	0.000	-1.821	0.730	-0.571	0.455	0.200	0.070	0.070	0.000	0.879
9	1.000	2.650	-2.650	0.000	-2.581	1.249	-1.131	0.529	0.200	0.110	-0.110	0.000	0.808
10	1.000	0.640	-0.640	0.000	-2.800	0.844	-0.590	0.605	0.200	0.030	0.030	0.000	0.462
11	1.000	0.850	-0.850	0.000	-1.882	0.722	-0.472	0.546	0.200	0.030	0.030	0.000	0.603
12	1.000	0.710	-0.710	0.000	-3.024	1.188	-0.894	0.782	0.200	0.090	-0.090	0.000	0.487
13	1.000	0.510	-0.510	0.000	-2.640	1.051	-0.660	0.818	0.200	0.010	0.010	0.000	0.477
14	1.000	0.780	-0.780	0.000	-1.603	0.786	-0.603	0.503	0.200	0.110	-0.110	0.000	0.465
15	1.000	0.230	0.230	0.000	-1.543	0.731	0.578	0.449	0.200	0.050	0.050	0.000	0.380
16	1.000	0.180	-0.180	0.000	-0.740	0.437	0.190	0.394	0.200	0.050	0.050	0.000	0.312
17	1.000	0.980	-0.980	0.000	-3.461	1.138	-1.021	0.502	0.200	0.020	-0.020	0.000	0.470
18	1.000	0.800	-0.800	0.000	-2.064	0.763	-0.524	0.555	0.200	0.030	0.030	0.000	0.598
19	1.000	0.270	-0.270	0.000	-0.783	0.438	-0.203	0.388	0.200	0.110	-0.110	0.000	0.263
20	1.000	0.740	0.740	0.000	-0.920	2.938	2.900	0.469	0.200	0.000	0.000	0.000	0.952
21	1.000	0.280	-0.280	0.000	-0.138	0.497	0.302	0.395	0.200	0.080	0.080	0.000	0.445
22	1.000	0.690	-0.690	0.000	-1.373	0.478	-0.143	0.456	0.200	0.080	0.080	0.000	0.572
23	1.000	0.390	-0.390	0.000	0.480	0.483	0.220	0.431	0.200	0.040	-0.040	0.000	0.409
24	1.000	0.030	0.030	0.000	-1.445	0.707	0.425	0.565	0.200	0.040	0.040	0.000	0.390
25	1.000	1.220	-1.220	0.000	-2.632	1.137	-0.962	0.606	0.200	0.010	0.010	0.000	0.790
26	1.000	0.010	0.010	0.000	-0.746	0.409	0.134	0.387	0.200	0.020	-0.020	0.000	0.280
27	1.000	0.660	-0.660	0.000	-1.273	0.505	-0.193	0.467	0.200	0.040	0.040	0.000	0.558
28	1.000	0.100	0.100	0.000	1.002	0.746	0.542	0.513	0.200	0.060	0.060	0.000	0.422
29	1.000	1.560	-1.560	0.000	-2.935	1.182	-0.975	0.667	0.200	0.060	0.060	0.000	0.830
30	1.000	0.540	-0.540	0.000	-1.029	0.419	0.041	0.417	0.200	0.110	0.110	0.000	0.540
31	1.000	0.930	-0.930	0.000	-0.607	0.460	0.113	0.446	0.200	0.140	0.140	0.000	0.746

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.000	0.970	-0.970	0.000	-0.315	0.387	-0.085	0.378	0.200	0.070	-0.070	0.000	0.409
33	1.000	0.680	-0.680	0.000	-0.326	0.519	0.214	0.472	0.200	0.130	0.130	0.000	0.670
34	1.000	1.610	-1.610	0.000	-2.747	1.164	-0.997	0.600	0.200	0.030	0.030	0.000	0.855
35	1.000	1.460	-1.460	0.000	-2.741	1.201	-0.931	0.760	0.200	0.050	0.050	0.000	0.831
36	1.000	1.130	-1.130	0.000	2.737	1.486	1.297	0.725	0.200	0.040	0.040	0.000	0.960
37	1.000	0.490	0.490	0.000	-0.153	0.812	0.627	0.517	0.200	0.010	-0.010	0.000	0.626
38	1.000	0.020	-0.020	0.000	-0.865	0.467	0.195	0.424	0.200	0.010	0.010	0.000	0.319
39	1.000	0.040	0.040	0.000	1.918	0.645	0.318	0.561	0.200	0.030	-0.030	0.000	0.427
40	1.000	0.350	-0.350	0.000	-0.305	0.371	-0.025	0.370	0.200	0.080	-0.080	0.000	0.327
41	1.000	0.150	0.150	0.000	-1.680	0.616	0.270	0.554	0.200	0.020	0.020	0.000	0.412
42	1.000	1.040	-1.040	0.000	-1.845	0.796	-0.605	0.517	0.200	0.030	-0.030	0.000	0.648
43	1.000	0.150	-0.150	0.000	0.523	0.382	-0.097	0.369	0.200	0.140	-0.140	0.000	0.426
44	1.000	0.400	0.400	0.000	-1.274	0.915	0.756	0.516	0.200	0.000	0.000	0.000	0.525
45	1.000	1.800	-1.800	0.000	-2.549	1.244	-1.119	0.543	0.200	0.090	-0.090	0.000	0.761
46	1.000	0.700	-0.700	0.000	0.383	0.608	0.463	0.394	0.200	0.120	0.120	0.000	0.672
47	1.000	0.200	0.200	0.000	2.528	0.978	0.688	0.695	0.200	0.080	0.080	0.000	0.421
48	1.000	0.100	0.100	0.000	-0.077	0.643	0.494	0.412	0.200	0.090	0.090	0.000	0.315
49	1.000	0.210	0.210	0.000	0.124	0.624	0.454	0.428	0.200	0.070	0.070	0.000	0.280
50	1.000	0.690	0.690	0.000	1.041	0.672	-0.399	0.540	0.200	0.010	-0.010	0.000	0.784

Table E.34. Item-Level Evaluation Criteria Summary Statistics, Condition 3, Sample Size = 50

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.311	0.566	0.071	0.562	-1.505	2.188	0.245	2.174	0.200	0.070	0.070	0.000	0.729
2	0.701	0.540	-0.389	0.374	-1.070	1.353	-0.780	1.105	0.200	0.130	-0.130	0.000	0.475
3	1.365	0.636	-0.005	0.636	-1.294	4.672	0.586	4.635	0.200	0.060	0.060	0.000	0.728
4	1.459	0.658	0.189	0.631	-1.514	2.574	0.436	2.537	0.200	0.050	0.050	0.000	0.791
5	0.793	0.577	0.003	0.577	-0.689	7.520	2.251	7.176	0.200	0.020	0.020	0.000	1.294
6	1.133	0.763	-0.127	0.752	0.346	10.438	2.876	10.034	0.200	0.020	0.020	0.000	1.062
7	1.533	0.522	-0.137	0.503	-1.409	0.725	-0.029	0.724	0.200	0.060	0.060	0.000	0.413
8	1.751	1.067	-0.969	0.446	-1.064	2.232	0.186	2.225	0.200	0.070	0.070	0.000	0.601
9	1.890	1.784	-1.760	0.290	-1.857	0.591	-0.407	0.429	0.200	0.110	-0.110	0.000	0.466
10	1.360	0.732	-0.280	0.677	-1.967	3.857	0.243	3.850	0.200	0.030	0.030	0.000	0.723
11	1.686	0.494	-0.164	0.466	-1.496	0.721	-0.086	0.716	0.200	0.030	0.030	0.000	0.450
12	1.498	0.661	-0.212	0.626	-1.325	5.819	0.805	5.763	0.200	0.090	-0.090	0.000	0.937
13	1.417	0.643	-0.093	0.636	-1.123	6.401	0.857	6.343	0.200	0.010	0.010	0.000	0.826
14	1.284	0.763	-0.496	0.580	-1.830	1.780	-0.830	1.574	0.200	0.110	-0.110	0.000	0.618
15	0.790	0.540	0.020	0.539	-0.891	4.968	1.229	4.813	0.200	0.050	0.050	0.000	0.915
16	1.050	0.497	-0.131	0.480	-0.887	0.619	0.043	0.617	0.200	0.050	0.050	0.000	0.411
17	1.592	0.721	-0.388	0.608	-2.455	3.279	-0.015	3.279	0.200	0.020	-0.020	0.000	0.591
18	1.610	0.562	-0.190	0.528	-1.440	1.911	0.100	1.908	0.200	0.030	0.030	0.000	0.593
19	0.887	0.605	-0.384	0.468	-1.433	1.964	-0.853	1.770	0.200	0.110	-0.110	0.000	0.563
20	0.345	0.230	0.085	0.214	-2.404	4.824	1.417	4.612	0.200	0.000	0.000	0.000	0.595
21	1.092	0.536	-0.188	0.502	-0.226	0.605	0.214	0.566	0.200	0.080	0.080	0.000	0.592
22	1.482	0.603	-0.208	0.566	-1.295	0.920	-0.065	0.917	0.200	0.080	0.080	0.000	0.511
23	0.843	0.680	-0.548	0.404	0.439	0.666	0.179	0.641	0.200	0.040	-0.040	0.000	0.581
24	1.049	0.559	0.079	0.554	-0.616	5.430	1.254	5.283	0.200	0.040	0.040	0.000	0.760
25	1.690	0.769	-0.530	0.557	-0.704	5.992	0.966	5.913	0.200	0.010	0.010	0.000	0.760
26	0.829	0.510	-0.161	0.484	-1.131	2.233	-0.251	2.219	0.200	0.020	-0.020	0.000	0.629
27	1.270	0.742	-0.390	0.632	-0.845	3.077	0.235	3.068	0.200	0.040	0.040	0.000	0.838
28	0.706	0.378	-0.194	0.325	1.332	1.280	0.872	0.936	0.200	0.060	0.060	0.000	0.647
29	1.889	0.763	-0.671	0.363	-1.533	2.752	0.427	2.719	0.200	0.060	0.060	0.000	0.726
30	1.436	0.498	-0.104	0.487	-0.873	0.465	0.197	0.421	0.200	0.110	0.110	0.000	0.432
31	1.589	0.594	-0.341	0.486	-0.479	0.465	0.241	0.397	0.200	0.140	0.140	0.000	0.643

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.306	0.834	-0.664	0.504	-0.341	0.577	-0.111	0.567	0.200	0.070	-0.070	0.000	0.509
33	1.310	0.648	-0.370	0.532	-0.306	0.529	0.234	0.474	0.200	0.130	0.130	0.000	0.685
34	1.850	0.843	-0.760	0.364	-2.092	0.933	-0.342	0.868	0.200	0.030	0.030	0.000	0.511
35	1.730	0.872	-0.730	0.476	-1.647	3.692	0.163	3.688	0.200	0.050	0.050	0.000	0.788
36	0.964	1.301	-1.166	0.576	4.272	5.067	2.832	4.202	0.200	0.040	0.040	0.000	1.042
37	0.457	0.335	-0.053	0.331	-0.970	1.589	-0.190	1.578	0.200	0.010	-0.010	0.000	0.496
38	0.943	0.555	-0.077	0.550	-1.330	1.145	-0.270	1.113	0.200	0.010	0.010	0.000	0.575
39	0.543	0.535	-0.417	0.335	4.325	4.536	2.725	3.626	0.200	0.030	-0.030	0.000	0.651
40	0.981	0.574	-0.369	0.440	-0.561	1.044	-0.281	1.005	0.200	0.080	-0.080	0.000	0.516
41	0.803	0.488	-0.047	0.486	-0.638	5.971	1.312	5.825	0.200	0.020	0.020	0.000	0.920
42	1.501	0.802	-0.539	0.593	-0.691	4.910	0.550	4.879	0.200	0.030	-0.030	0.000	0.792
43	0.677	0.565	-0.473	0.308	0.641	0.708	0.021	0.707	0.200	0.140	-0.140	0.000	0.473
44	0.562	0.370	-0.038	0.368	-2.804	2.913	-0.774	2.808	0.200	0.000	0.000	0.000	0.434
45	1.798	1.093	-1.003	0.434	-2.017	1.050	-0.587	0.870	0.200	0.090	-0.090	0.000	0.521
46	1.325	0.609	-0.375	0.479	0.365	0.558	0.445	0.338	0.200	0.120	0.120	0.000	0.682
47	0.776	0.449	-0.024	0.449	4.051	3.955	2.211	3.279	0.200	0.080	0.080	0.000	0.589
48	0.763	0.453	-0.137	0.431	-0.413	1.272	0.157	1.262	0.200	0.090	0.090	0.000	0.588
49	0.776	0.405	-0.014	0.404	-0.055	0.690	0.275	0.633	0.200	0.070	0.070	0.000	0.487
50	0.398	0.276	0.088	0.262	2.499	2.819	1.059	2.613	0.200	0.010	-0.010	0.000	0.484

Table E.35. Item-Level Evaluation Criteria Summary Statistics, Condition 4, Sample Size = 50

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.312	0.566	0.072	0.561	-1.423	1.063	0.327	1.012	0.200	0.070	0.070	0.000	0.528
2	0.701	0.540	-0.389	0.374	-0.763	0.823	-0.473	0.673	0.200	0.130	-0.130	0.000	0.443
3	1.365	0.636	-0.006	0.636	-1.732	1.706	0.148	1.700	0.200	0.060	0.060	0.000	0.543
4	1.460	0.658	0.190	0.630	-1.618	1.144	0.333	1.094	0.200	0.050	0.050	0.000	0.563
5	0.792	0.577	0.002	0.577	-2.431	1.581	0.509	1.497	0.200	0.020	0.020	0.000	0.763
6	1.133	0.763	-0.128	0.752	-2.837	2.347	-0.307	2.327	0.200	0.020	0.020	0.000	0.716
7	1.533	0.522	-0.137	0.504	-1.087	0.528	0.293	0.440	0.200	0.060	0.060	0.000	0.406
8	1.751	1.067	-0.969	0.446	-1.174	1.070	0.076	1.068	0.200	0.070	0.070	0.000	0.517
9	1.890	1.784	-1.760	0.291	-1.360	0.293	0.091	0.279	0.200	0.110	-0.110	0.000	0.412
10	1.359	0.733	-0.281	0.677	-2.232	1.560	-0.022	1.560	0.200	0.030	0.030	0.000	0.612
11	1.686	0.494	-0.164	0.466	-1.140	0.524	0.270	0.449	0.200	0.030	0.030	0.000	0.437
12	1.498	0.661	-0.212	0.626	-2.158	1.863	-0.028	1.863	0.200	0.090	-0.090	0.000	0.672
13	1.417	0.643	-0.093	0.636	-2.050	2.019	-0.070	2.017	0.200	0.010	0.010	0.000	0.645
14	1.284	0.763	-0.496	0.580	-1.329	0.977	-0.329	0.920	0.200	0.110	-0.110	0.000	0.521
15	0.790	0.540	0.020	0.539	-1.786	1.376	0.334	1.335	0.200	0.050	0.050	0.000	0.633
16	1.050	0.497	-0.130	0.480	-0.725	0.436	0.205	0.385	0.200	0.050	0.050	0.000	0.371
17	1.592	0.721	-0.388	0.608	-2.287	1.289	0.153	1.279	0.200	0.020	-0.020	0.000	0.643
18	1.610	0.561	-0.190	0.528	-1.365	0.919	0.175	0.902	0.200	0.030	0.030	0.000	0.479
19	0.887	0.605	-0.384	0.468	-0.950	0.938	-0.370	0.862	0.200	0.110	-0.110	0.000	0.536
20	0.345	0.231	0.085	0.214	-1.839	2.301	1.981	1.172	0.200	0.000	0.000	0.000	0.572
21	1.093	0.536	-0.187	0.502	-0.273	0.398	0.167	0.361	0.200	0.080	0.080	0.000	0.527
22	1.482	0.603	-0.208	0.566	-1.002	0.600	0.228	0.555	0.200	0.080	0.080	0.000	0.478
23	0.843	0.680	-0.547	0.404	0.193	0.424	-0.067	0.419	0.200	0.040	-0.040	0.000	0.484
24	1.049	0.559	0.079	0.554	-1.451	1.595	0.419	1.539	0.200	0.040	0.040	0.000	0.624
25	1.690	0.770	-0.530	0.558	-1.795	1.729	-0.125	1.724	0.200	0.010	0.010	0.000	0.501
26	0.829	0.510	-0.161	0.484	-0.902	0.656	-0.022	0.655	0.200	0.020	-0.020	0.000	0.522
27	1.270	0.742	-0.390	0.632	-1.143	1.013	-0.063	1.011	0.200	0.040	0.040	0.000	0.615
28	0.706	0.379	-0.194	0.326	0.783	0.720	0.323	0.643	0.200	0.060	0.060	0.000	0.550
29	1.889	0.763	-0.672	0.363	-1.676	1.156	0.284	1.121	0.200	0.060	0.060	0.000	0.591
30	1.436	0.498	-0.104	0.487	-0.734	0.435	0.336	0.277	0.200	0.110	0.110	0.000	0.400
31	1.590	0.594	-0.341	0.486	-0.463	0.378	0.257	0.277	0.200	0.140	0.140	0.000	0.598

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.306	0.834	-0.664	0.504	-0.363	0.376	-0.133	0.352	0.200	0.070	-0.070	0.000	0.461
33	1.310	0.648	-0.370	0.532	-0.346	0.372	0.194	0.318	0.200	0.130	0.130	0.000	0.642
34	1.850	0.843	-0.760	0.364	-1.507	0.580	0.243	0.526	0.200	0.030	0.030	0.000	0.481
35	1.730	0.872	-0.730	0.476	-1.747	1.751	0.063	1.750	0.200	0.050	0.050	0.000	0.601
36	0.964	1.301	-1.166	0.576	2.035	1.280	0.595	1.134	0.200	0.040	0.040	0.000	0.833
37	0.457	0.336	-0.053	0.331	-0.549	0.850	0.231	0.818	0.200	0.010	-0.010	0.000	0.463
38	0.943	0.555	-0.077	0.550	-0.989	0.677	0.072	0.674	0.200	0.010	0.010	0.000	0.505
39	0.544	0.535	-0.416	0.336	1.869	1.311	0.269	1.283	0.200	0.030	-0.030	0.000	0.578
40	0.981	0.575	-0.369	0.440	-0.488	0.649	-0.208	0.615	0.200	0.080	-0.080	0.000	0.451
41	0.803	0.488	-0.047	0.486	-1.998	1.822	-0.048	1.821	0.200	0.020	0.020	0.000	0.546
42	1.501	0.802	-0.539	0.593	-1.202	0.518	0.038	0.517	0.200	0.030	-0.030	0.000	0.536
43	0.678	0.565	-0.473	0.309	0.348	0.595	-0.272	0.529	0.200	0.140	-0.140	0.000	0.383
44	0.562	0.371	-0.038	0.369	-1.919	1.292	0.112	1.287	0.200	0.000	0.000	0.000	0.465
45	1.798	1.092	-1.002	0.434	-1.445	0.531	-0.015	0.531	0.200	0.090	-0.090	0.000	0.438
46	1.324	0.609	-0.376	0.479	0.117	0.314	0.197	0.245	0.200	0.120	0.120	0.000	0.646
47	0.776	0.449	-0.024	0.449	1.918	0.990	0.078	0.987	0.200	0.080	0.080	0.000	0.697
48	0.763	0.452	-0.137	0.431	-0.302	0.535	0.268	0.463	0.200	0.090	0.090	0.000	0.542
49	0.777	0.404	-0.013	0.404	-0.118	0.439	0.212	0.384	0.200	0.070	0.070	0.000	0.446
50	0.398	0.277	0.088	0.262	1.328	1.427	-0.112	1.423	0.200	0.010	-0.010	0.000	0.456

Table E.36. Item-Level Evaluation Criteria Summary Statistics, Condition 5, Sample Size = 50

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	0.900	0.458	-0.340	0.306	-2.135	1.289	-0.385	1.231	0.197	0.067	0.067	0.005	0.527
2	0.749	0.613	-0.341	0.510	-0.840	0.972	-0.550	0.801	0.198	0.133	-0.133	0.010	0.481
3	0.981	0.596	-0.389	0.451	-2.686	1.800	-0.806	1.609	0.197	0.057	0.057	0.006	0.587
4	0.986	0.480	-0.284	0.387	-2.454	1.308	-0.504	1.207	0.198	0.048	0.048	0.004	0.503
5	0.576	0.349	-0.214	0.275	-3.882	2.814	-0.942	2.651	0.200	0.020	0.020	0.000	0.503
6	0.741	0.634	-0.520	0.364	-3.064	4.474	-0.534	4.442	0.200	0.020	0.020	0.002	0.642
7	1.245	0.675	-0.425	0.524	-1.574	0.854	-0.194	0.832	0.195	0.056	0.055	0.011	0.536
8	1.662	1.242	-1.058	0.651	-1.488	0.912	-0.238	0.881	0.191	0.062	0.061	0.011	0.702
9	1.488	2.211	-2.162	0.463	-2.128	0.982	-0.678	0.710	0.197	0.113	-0.113	0.008	0.616
10	0.940	0.830	-0.700	0.445	-3.468	2.123	-1.258	1.710	0.199	0.029	0.029	0.004	0.556
11	1.361	0.690	-0.489	0.487	-1.660	0.954	-0.250	0.921	0.193	0.026	0.023	0.011	0.558
12	0.957	0.846	-0.753	0.386	-3.239	2.888	-1.109	2.667	0.199	0.091	-0.091	0.004	0.839
13	0.955	0.691	-0.555	0.412	-2.895	2.692	-0.915	2.532	0.198	0.010	0.008	0.005	0.795
14	1.054	0.852	-0.726	0.446	-1.807	1.401	-0.807	1.146	0.198	0.113	-0.112	0.010	0.605
15	0.637	0.382	-0.133	0.358	-2.793	1.677	-0.673	1.537	0.198	0.049	0.048	0.005	0.471
16	0.883	0.485	-0.297	0.383	-0.926	0.678	0.004	0.677	0.193	0.045	0.043	0.013	0.481
17	0.918	1.116	-1.062	0.342	-4.018	2.217	-1.578	1.557	0.200	0.020	-0.020	0.002	0.531
18	1.288	0.729	-0.512	0.519	-1.927	0.975	-0.387	0.895	0.196	0.027	0.026	0.008	0.550
19	0.808	0.572	-0.462	0.338	-1.140	1.114	-0.560	0.963	0.197	0.114	-0.113	0.009	0.519
20	0.359	0.194	0.099	0.167	-2.971	2.102	0.849	1.923	0.200	0.003	0.000	0.003	0.403
21	1.062	0.590	-0.218	0.548	-0.151	0.594	0.289	0.519	0.184	0.065	0.064	0.013	0.617
22	1.159	0.681	-0.531	0.426	-1.335	0.715	-0.105	0.708	0.192	0.073	0.072	0.011	0.581
23	0.915	0.648	-0.475	0.441	0.512	0.663	0.252	0.613	0.190	0.054	-0.051	0.019	0.605
24	0.826	0.408	-0.144	0.382	-2.157	1.840	-0.287	1.818	0.197	0.038	0.037	0.007	0.573
25	1.289	1.082	-0.931	0.552	-2.401	2.413	-0.731	2.299	0.198	0.010	0.008	0.007	0.738
26	0.812	0.476	-0.178	0.441	-1.191	1.121	-0.311	1.077	0.198	0.023	-0.022	0.008	0.539
27	1.213	0.816	-0.447	0.683	-1.457	1.246	-0.377	1.187	0.195	0.037	0.035	0.012	0.707
28	0.738	0.403	-0.162	0.369	1.349	1.214	0.889	0.826	0.184	0.047	0.044	0.016	0.676
29	1.270	1.361	-1.290	0.435	-2.815	1.760	-0.855	1.538	0.196	0.057	0.056	0.007	0.706
30	1.170	0.561	-0.371	0.421	-0.926	0.458	0.144	0.434	0.189	0.100	0.099	0.012	0.523
31	1.529	0.703	-0.401	0.577	-0.433	0.512	0.287	0.424	0.178	0.120	0.118	0.018	0.664

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.402	0.943	-0.568	0.753	-0.227	0.549	0.003	0.549	0.193	0.079	-0.077	0.017	0.541
33	1.198	0.651	-0.482	0.438	-0.251	0.557	0.289	0.476	0.184	0.115	0.114	0.018	0.687
34	1.352	1.334	-1.258	0.444	-2.481	1.345	-0.731	1.129	0.196	0.027	0.026	0.007	0.709
35	1.255	1.291	-1.205	0.463	-2.323	2.139	-0.513	2.076	0.196	0.046	0.046	0.007	0.903
36	0.734	1.449	-1.396	0.388	3.906	3.196	2.466	2.034	0.166	0.020	0.006	0.019	1.118
37	0.453	0.248	-0.058	0.241	-0.686	1.161	0.094	1.157	0.198	0.014	-0.012	0.008	0.445
38	0.819	0.470	-0.201	0.425	-1.329	1.272	-0.269	1.243	0.198	0.014	0.008	0.012	0.584
39	0.569	0.531	-0.391	0.360	3.377	2.463	1.777	1.706	0.188	0.045	-0.042	0.017	0.653
40	0.974	0.627	-0.377	0.501	-0.425	0.801	-0.145	0.788	0.196	0.086	-0.084	0.016	0.548
41	0.665	0.320	-0.185	0.262	-2.688	1.761	-0.738	1.599	0.199	0.020	0.019	0.005	0.503
42	1.238	0.979	-0.802	0.562	-1.611	1.990	-0.371	1.955	0.194	0.037	-0.036	0.009	0.828
43	0.737	0.624	-0.414	0.467	0.686	0.562	0.066	0.558	0.194	0.147	-0.146	0.011	0.518
44	0.521	0.257	-0.079	0.244	-2.634	1.652	-0.604	1.538	0.199	0.004	-0.001	0.004	0.395
45	1.470	1.409	-1.330	0.467	-2.220	1.301	-0.790	1.034	0.196	0.095	-0.094	0.007	0.621
46	1.342	0.610	-0.358	0.494	0.424	0.603	0.504	0.331	0.174	0.096	0.094	0.019	0.665
47	0.554	0.333	-0.247	0.223	3.990	2.855	2.150	1.878	0.174	0.057	0.054	0.018	0.596
48	0.700	0.364	-0.200	0.304	-0.245	0.824	0.325	0.758	0.192	0.082	0.082	0.010	0.564
49	0.708	0.365	-0.082	0.356	0.044	0.731	0.374	0.628	0.189	0.060	0.059	0.013	0.488
50	0.441	0.293	0.131	0.261	1.974	1.592	0.534	1.500	0.199	0.016	-0.011	0.011	0.491

Table E.37. Item-Level Evaluation Criteria Summary Statistics, Condition 0, Sample Size = 25

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	2.048	1.025	0.808	0.630	-1.170	0.778	0.580	0.518	0.197	0.067	0.067	0.005	0.268
2	1.456	1.845	0.366	1.808	-0.671	1.112	-0.381	1.044	0.197	0.134	-0.133	0.012	0.343
3	1.717	0.894	0.347	0.824	-1.752	0.909	0.128	0.899	0.197	0.057	0.057	0.005	0.249
4	2.171	1.439	0.901	1.122	-1.627	1.776	0.323	1.746	0.193	0.044	0.043	0.008	0.345
5	1.423	1.226	0.633	1.050	-2.653	2.026	0.287	2.006	0.198	0.019	0.018	0.006	0.371
6	1.785	1.061	0.525	0.922	-2.372	1.046	0.158	1.034	0.202	0.022	0.022	0.006	0.281
7	2.341	1.437	0.671	1.271	-1.217	0.593	0.163	0.570	0.195	0.056	0.055	0.008	0.267
8	3.224	1.213	0.504	1.104	-1.043	0.508	0.207	0.464	0.191	0.061	0.061	0.009	0.261
9	2.743	1.190	-0.907	0.771	-1.426	0.487	0.024	0.486	0.197	0.114	-0.113	0.006	0.299
10	1.695	0.944	0.055	0.943	-2.420	1.790	-0.210	1.777	0.199	0.029	0.029	0.003	0.256
11	2.468	1.311	0.618	1.157	-1.012	0.918	0.398	0.827	0.195	0.027	0.025	0.011	0.289
12	1.870	1.091	0.160	1.079	-2.930	2.339	-0.800	2.197	0.197	0.094	-0.093	0.006	0.256
13	1.953	0.812	0.443	0.681	-1.605	0.631	0.375	0.508	0.201	0.013	0.011	0.006	0.248
14	2.665	1.919	0.885	1.703	-1.524	1.873	-0.524	1.798	0.205	0.106	-0.105	0.017	0.382
15	1.058	0.746	0.288	0.688	-2.543	2.255	-0.423	2.215	0.198	0.048	0.048	0.004	0.373
16	1.436	0.833	0.256	0.792	-0.558	0.719	0.373	0.615	0.192	0.042	0.042	0.008	0.284
17	2.546	1.161	0.566	1.013	-2.060	1.240	0.380	1.180	0.203	0.018	-0.017	0.008	0.302
18	2.654	1.336	0.854	1.027	-1.039	0.762	0.501	0.575	0.197	0.029	0.027	0.010	0.318
19	1.318	0.748	0.048	0.746	-0.787	0.766	-0.207	0.738	0.199	0.112	-0.111	0.013	0.253
20	0.704	0.656	0.444	0.482	-1.815	2.636	2.005	1.711	0.200	0.006	0.000	0.006	0.392
21	2.478	2.123	1.198	1.753	-0.050	0.586	0.390	0.437	0.189	0.070	0.069	0.009	0.268
22	2.244	1.501	0.554	1.395	-0.996	0.879	0.234	0.847	0.190	0.071	0.070	0.012	0.358
23	1.466	0.774	0.076	0.770	0.685	1.056	0.425	0.967	0.193	0.050	-0.048	0.017	0.403
24	1.845	1.707	0.875	1.465	-1.697	1.787	0.173	1.779	0.197	0.038	0.037	0.009	0.417
25	2.882	1.400	0.662	1.234	-1.698	1.598	-0.028	1.598	0.203	0.016	0.013	0.009	0.337
26	1.209	0.681	0.219	0.645	-1.204	0.933	-0.324	0.875	0.198	0.023	-0.022	0.008	0.264
27	2.603	2.059	0.943	1.831	-1.365	1.651	-0.285	1.626	0.195	0.038	0.035	0.014	0.334
28	1.083	0.433	0.183	0.392	1.147	0.901	0.687	0.584	0.189	0.050	0.049	0.011	0.232
29	2.693	1.211	0.133	1.204	-1.783	1.495	0.178	1.484	0.197	0.057	0.057	0.008	0.425
30	2.366	1.654	0.826	1.433	-0.798	0.594	0.272	0.528	0.189	0.100	0.099	0.013	0.278
31	2.420	1.721	0.490	1.649	-0.253	0.733	0.468	0.564	0.187	0.127	0.127	0.012	0.319

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	2.395	2.013	0.425	1.968	-0.212	0.383	0.018	0.382	0.195	0.076	-0.075	0.014	0.275
33	1.607	0.790	-0.073	0.786	-0.185	0.596	0.355	0.478	0.197	0.127	0.127	0.012	0.318
34	2.935	1.090	0.325	1.041	-1.506	0.683	0.244	0.638	0.192	0.023	0.022	0.008	0.249
35	2.510	1.065	0.050	1.064	-1.746	1.245	0.064	1.243	0.195	0.046	0.045	0.008	0.255
36	1.252	1.153	-0.878	0.747	2.423	1.685	0.983	1.368	0.183	0.031	0.023	0.020	0.432
37	0.776	0.627	0.266	0.568	0.081	1.129	0.861	0.731	0.196	0.016	-0.014	0.008	0.261
38	1.236	0.920	0.216	0.894	-1.861	1.956	-0.801	1.784	0.199	0.011	0.009	0.006	0.333
39	0.793	0.632	-0.167	0.610	2.938	3.175	1.338	2.879	0.198	0.033	-0.033	0.006	0.344
40	2.226	2.155	0.876	1.969	-0.317	0.389	-0.037	0.387	0.193	0.089	-0.088	0.015	0.260
41	1.790	1.350	0.940	0.970	-1.401	1.298	0.549	1.176	0.199	0.021	0.019	0.010	0.401
42	2.624	1.200	0.584	1.048	-1.143	0.734	0.097	0.728	0.193	0.038	-0.037	0.009	0.354
43	1.119	1.062	-0.031	1.062	0.625	0.499	0.005	0.499	0.194	0.146	-0.146	0.012	0.269
44	0.950	0.927	0.350	0.858	-2.561	1.901	-0.531	1.826	0.199	0.003	-0.001	0.003	0.248
45	2.795	0.705	-0.005	0.705	-1.548	0.493	-0.118	0.479	0.198	0.092	-0.092	0.009	0.226
46	2.824	1.978	1.124	1.628	0.239	0.413	0.319	0.262	0.183	0.104	0.103	0.017	0.356
47	1.152	0.980	0.352	0.915	2.688	1.653	0.848	1.418	0.189	0.073	0.069	0.024	0.309
48	1.199	0.862	0.299	0.808	-0.245	0.934	0.325	0.876	0.193	0.083	0.083	0.009	0.273
49	1.609	1.320	0.819	1.036	0.413	1.008	0.743	0.681	0.186	0.057	0.056	0.009	0.329
50	0.658	0.569	0.348	0.450	1.528	1.101	0.088	1.097	0.195	0.017	-0.015	0.009	0.317

Table E.38. Item-Level Evaluation Criteria Summary Statistics, Condition 1, Sample Size = 25

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.653	0.573	0.413	0.397	-1.371	0.693	0.379	0.581	0.194	0.064	0.064	0.005	0.221
2	0.789	0.614	-0.301	0.535	-1.203	2.009	-0.913	1.789	0.195	0.136	-0.135	0.011	0.342
3	1.524	0.586	0.154	0.565	-2.057	1.500	-0.177	1.490	0.198	0.058	0.058	0.004	0.266
4	1.637	0.689	0.367	0.583	-0.080	4.348	1.870	3.925	0.193	0.045	0.043	0.010	0.331
5	1.152	0.797	0.362	0.710	0.646	7.099	3.586	6.127	0.200	0.021	0.020	0.006	0.673
6	1.368	0.717	0.108	0.709	-3.077	1.652	-0.547	1.559	0.200	0.020	0.020	0.004	0.196
7	1.718	0.506	0.048	0.504	-1.380	0.676	0.000	0.676	0.195	0.056	0.055	0.009	0.293
8	1.910	0.843	-0.810	0.234	-1.318	0.554	-0.068	0.550	0.189	0.060	0.059	0.006	0.274
9	1.882	1.785	-1.768	0.244	-1.713	0.595	-0.263	0.534	0.194	0.116	-0.116	0.005	0.245
10	1.370	0.664	-0.270	0.607	-2.563	1.483	-0.353	1.441	0.198	0.029	0.028	0.004	0.226
11	1.620	0.632	-0.230	0.589	-1.307	0.954	0.103	0.948	0.193	0.024	0.023	0.008	0.248
12	1.444	0.801	-0.266	0.756	0.783	8.441	2.913	7.923	0.198	0.093	-0.093	0.008	0.524
13	1.755	0.509	0.245	0.447	-1.749	0.595	0.231	0.548	0.200	0.012	0.010	0.007	0.216
14	1.438	0.755	-0.342	0.674	-0.551	4.035	0.449	4.010	0.199	0.111	-0.111	0.011	0.431
15	0.944	0.688	0.174	0.666	1.522	7.402	3.642	6.444	0.198	0.048	0.048	0.004	0.681
16	1.242	0.504	0.062	0.500	-0.553	0.802	0.377	0.708	0.193	0.044	0.043	0.009	0.278
17	1.747	0.625	-0.233	0.580	-0.938	4.217	1.503	3.940	0.199	0.021	-0.021	0.005	0.512
18	1.743	0.507	-0.057	0.504	-1.403	1.067	0.138	1.058	0.193	0.025	0.023	0.009	0.243
19	1.058	0.679	-0.212	0.645	-1.193	1.339	-0.613	1.191	0.196	0.114	-0.114	0.006	0.297
20	0.573	0.519	0.313	0.415	0.400	7.180	4.220	5.809	0.199	0.008	-0.001	0.008	0.540
21	1.445	0.563	0.165	0.538	-0.124	0.671	0.316	0.591	0.187	0.067	0.067	0.009	0.290
22	1.493	0.654	-0.198	0.623	-1.253	1.141	-0.023	1.140	0.190	0.071	0.070	0.010	0.342
23	0.952	0.654	-0.438	0.486	0.820	1.426	0.560	1.311	0.189	0.052	-0.051	0.010	0.450
24	1.258	0.835	0.288	0.784	-0.441	5.037	1.429	4.830	0.193	0.034	0.033	0.008	0.515
25	1.768	0.706	-0.453	0.543	0.769	8.138	2.439	7.764	0.198	0.012	0.008	0.008	0.443
26	1.062	0.674	0.072	0.671	-0.238	3.667	0.642	3.610	0.198	0.024	-0.022	0.009	0.383
27	1.486	0.676	-0.174	0.653	0.065	4.208	1.145	4.049	0.194	0.036	0.034	0.011	0.346
28	0.775	0.499	-0.125	0.483	1.624	1.523	1.164	0.982	0.190	0.051	0.050	0.010	0.330
29	1.740	0.970	-0.820	0.519	-0.553	3.853	1.408	3.586	0.195	0.055	0.055	0.007	0.490
30	1.553	0.451	0.013	0.451	-0.893	0.542	0.177	0.512	0.190	0.101	0.100	0.011	0.248
31	1.428	0.792	-0.502	0.612	-0.318	0.694	0.403	0.565	0.186	0.126	0.126	0.010	0.330

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.358	0.929	-0.613	0.698	-0.388	0.634	-0.158	0.614	0.190	0.081	-0.080	0.011	0.325
33	1.120	0.764	-0.560	0.519	-0.242	0.754	0.298	0.693	0.194	0.125	0.124	0.009	0.404
34	1.868	0.863	-0.743	0.440	-1.882	0.758	-0.132	0.746	0.192	0.022	0.022	0.006	0.194
35	1.828	0.820	-0.632	0.523	0.241	6.461	2.051	6.127	0.196	0.047	0.046	0.009	0.280
36	1.060	1.286	-1.070	0.714	3.437	3.306	1.997	2.635	0.190	0.037	0.030	0.022	0.484
37	0.643	0.542	0.133	0.526	0.233	1.705	1.013	1.372	0.195	0.017	-0.015	0.008	0.256
38	0.955	0.689	-0.065	0.686	0.180	6.561	1.240	6.443	0.198	0.010	0.008	0.006	0.471
39	0.603	0.652	-0.358	0.546	3.025	4.384	1.425	4.146	0.196	0.035	-0.034	0.006	0.500
40	1.271	0.699	-0.079	0.695	-0.517	0.663	-0.237	0.619	0.189	0.091	-0.091	0.010	0.272
41	1.264	0.772	0.414	0.651	-0.378	4.262	1.572	3.961	0.198	0.019	0.018	0.007	0.537
42	1.723	0.582	-0.318	0.488	-1.395	0.837	-0.155	0.823	0.193	0.038	-0.038	0.007	0.324
43	0.729	0.629	-0.421	0.467	0.738	0.763	0.118	0.754	0.192	0.149	-0.148	0.009	0.300
44	0.715	0.640	0.115	0.630	-1.603	4.754	0.428	4.735	0.199	0.003	-0.001	0.003	0.261
45	1.891	0.979	-0.909	0.362	-1.933	0.766	-0.503	0.578	0.195	0.095	-0.095	0.007	0.211
46	1.499	0.716	-0.201	0.687	0.243	0.443	0.323	0.302	0.177	0.098	0.097	0.014	0.429
47	0.954	0.760	0.154	0.744	5.461	6.869	3.621	5.837	0.192	0.076	0.072	0.024	0.326
48	0.909	0.597	0.009	0.597	-0.753	2.343	-0.183	2.336	0.192	0.082	0.082	0.007	0.297
49	1.182	0.660	0.392	0.531	0.462	1.216	0.792	0.923	0.184	0.055	0.054	0.010	0.332
50	0.473	0.448	0.163	0.418	2.953	3.544	1.513	3.205	0.194	0.018	-0.016	0.009	0.313

Table E.39. Item-Level Evaluation Criteria Summary Statistics, Condition 2, Sample Size = 25

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.000	0.240	-0.240	0.000	-1.766	0.751	-0.016	0.750	0.200	0.070	0.070	0.000	0.203
2	1.000	0.090	-0.090	0.000	-0.258	0.694	0.032	0.693	0.200	0.130	-0.130	0.000	0.292
3	1.000	0.370	-0.370	0.000	-2.103	0.549	-0.223	0.501	0.200	0.060	0.060	0.000	0.181
4	1.000	0.270	-0.270	0.000	-1.781	0.772	0.169	0.753	0.200	0.050	0.050	0.000	0.233
5	1.000	0.210	0.210	0.000	-2.176	1.088	0.764	0.774	0.200	0.020	0.020	0.000	0.227
6	1.000	0.260	-0.260	0.000	-2.699	0.547	-0.169	0.520	0.200	0.020	0.020	0.000	0.132
7	1.000	0.670	-0.670	0.000	-1.673	0.576	-0.293	0.496	0.200	0.060	0.060	0.000	0.291
8	1.000	1.720	-1.720	0.000	-1.838	0.841	-0.588	0.601	0.200	0.070	0.070	0.000	0.422
9	1.000	2.650	-2.650	0.000	-2.365	1.075	-0.915	0.564	0.200	0.110	-0.110	0.000	0.320
10	1.000	0.640	-0.640	0.000	-2.491	0.593	-0.281	0.522	0.200	0.030	0.030	0.000	0.174
11	1.000	0.850	-0.850	0.000	-1.481	0.666	-0.071	0.662	0.200	0.030	0.030	0.000	0.277
12	1.000	0.710	-0.710	0.000	-2.940	1.190	-0.810	0.872	0.200	0.090	-0.090	0.000	0.220
13	1.000	0.510	-0.510	0.000	-2.258	0.398	-0.278	0.286	0.200	0.010	0.010	0.000	0.148
14	1.000	0.780	-0.780	0.000	-1.694	1.088	-0.694	0.838	0.200	0.110	-0.110	0.000	0.242
15	1.000	0.230	0.230	0.000	-1.487	1.093	0.633	0.891	0.200	0.050	0.050	0.000	0.304
16	1.000	0.180	-0.180	0.000	-0.524	0.815	0.406	0.707	0.200	0.050	0.050	0.000	0.226
17	1.000	0.980	-0.980	0.000	-2.861	0.588	-0.421	0.410	0.200	0.020	-0.020	0.000	0.168
18	1.000	0.800	-0.800	0.000	-1.611	0.350	-0.071	0.342	0.200	0.030	0.030	0.000	0.211
19	1.000	0.270	-0.270	0.000	-0.663	0.420	-0.083	0.411	0.200	0.110	-0.110	0.000	0.191
20	1.000	0.740	0.740	0.000	-0.688	3.223	3.132	0.760	0.200	0.000	0.000	0.000	0.501
21	1.000	0.280	-0.280	0.000	-0.053	0.677	0.388	0.556	0.200	0.080	0.080	0.000	0.241
22	1.000	0.690	-0.690	0.000	-1.203	0.609	0.028	0.608	0.200	0.080	0.080	0.000	0.308
23	1.000	0.390	-0.390	0.000	0.878	1.018	0.618	0.809	0.200	0.040	-0.040	0.000	0.404
24	1.000	0.030	0.030	0.000	-1.262	0.813	0.608	0.539	0.200	0.040	0.040	0.000	0.243
25	1.000	1.220	-1.220	0.000	-2.303	0.885	-0.633	0.618	0.200	0.010	0.010	0.000	0.337
26	1.000	0.010	0.010	0.000	-1.100	0.675	-0.220	0.639	0.200	0.020	-0.020	0.000	0.204
27	1.000	0.660	-0.660	0.000	-1.606	1.148	-0.526	1.021	0.200	0.040	0.040	0.000	0.353
28	1.000	0.100	0.100	0.000	1.410	1.217	0.950	0.761	0.200	0.060	0.060	0.000	0.252
29	1.000	1.560	-1.560	0.000	-2.364	0.776	-0.404	0.662	0.200	0.060	0.060	0.000	0.323
30	1.000	0.540	-0.540	0.000	-1.088	0.571	-0.018	0.570	0.200	0.110	0.110	0.000	0.290
31	1.000	0.930	-0.930	0.000	-0.361	0.642	0.359	0.532	0.200	0.140	0.140	0.000	0.375

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.000	0.970	-0.970	0.000	-0.283	0.569	-0.053	0.566	0.200	0.070	-0.070	0.000	0.274
33	1.000	0.680	-0.680	0.000	-0.190	0.700	0.350	0.606	0.200	0.130	0.130	0.000	0.373
34	1.000	1.610	-1.610	0.000	-2.330	0.742	-0.580	0.462	0.200	0.030	0.030	0.000	0.347
35	1.000	1.460	-1.460	0.000	-2.278	0.767	-0.468	0.607	0.200	0.050	0.050	0.000	0.367
36	1.000	1.130	-1.130	0.000	2.720	1.987	1.280	1.520	0.200	0.040	0.040	0.000	0.423
37	1.000	0.490	0.490	0.000	0.263	1.158	1.043	0.504	0.200	0.010	-0.010	0.000	0.364
38	1.000	0.020	-0.020	0.000	-1.388	0.852	-0.328	0.787	0.200	0.010	0.010	0.000	0.257
39	1.000	0.040	0.040	0.000	1.113	2.247	-0.487	2.194	0.200	0.030	-0.030	0.000	0.436
40	1.000	0.350	-0.350	0.000	-0.254	0.462	0.026	0.462	0.200	0.080	-0.080	0.000	0.175
41	1.000	0.150	0.150	0.000	-1.430	0.846	0.520	0.667	0.200	0.020	0.020	0.000	0.235
42	1.000	1.040	-1.040	0.000	-1.760	0.967	-0.520	0.816	0.200	0.030	-0.030	0.000	0.346
43	1.000	0.150	-0.150	0.000	0.652	0.453	0.032	0.452	0.200	0.140	-0.140	0.000	0.270
44	1.000	0.400	0.400	0.000	-1.437	0.891	0.593	0.665	0.200	0.000	0.000	0.000	0.241
45	1.000	1.800	-1.800	0.000	-2.553	1.214	-1.123	0.463	0.200	0.090	-0.090	0.000	0.333
46	1.000	0.700	-0.700	0.000	0.403	0.714	0.483	0.525	0.200	0.120	0.120	0.000	0.348
47	1.000	0.200	0.200	0.000	2.848	2.165	1.008	1.917	0.200	0.080	0.080	0.000	0.337
48	1.000	0.100	0.100	0.000	0.054	0.826	0.624	0.541	0.200	0.090	0.090	0.000	0.201
49	1.000	0.210	0.210	0.000	0.613	1.174	0.943	0.699	0.200	0.070	0.070	0.000	0.272
50	1.000	0.690	0.690	0.000	1.188	0.604	-0.252	0.549	0.200	0.010	-0.010	0.000	0.388

Table E.40. Item-Level Evaluation Criteria Summary Statistics, Condition 3, Sample Size = 25

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.653	0.573	0.413	0.397	-1.370	0.698	0.380	0.586	0.200	0.070	0.070	0.000	0.217
2	0.789	0.614	-0.301	0.535	-1.197	2.013	-0.907	1.797	0.200	0.130	-0.130	0.000	0.335
3	1.524	0.586	0.154	0.565	-2.059	1.500	-0.179	1.489	0.200	0.060	0.060	0.000	0.265
4	1.637	0.689	0.367	0.583	-0.083	4.351	1.868	3.930	0.200	0.050	0.050	0.000	0.323
5	1.152	0.797	0.362	0.710	0.645	7.111	3.585	6.141	0.200	0.020	0.020	0.000	0.672
6	1.368	0.717	0.108	0.709	-3.081	1.656	-0.551	1.562	0.200	0.020	0.020	0.000	0.195
7	1.718	0.506	0.048	0.504	-1.388	0.673	-0.008	0.673	0.200	0.060	0.060	0.000	0.291
8	1.910	0.843	-0.810	0.234	-1.325	0.562	-0.075	0.557	0.200	0.070	0.070	0.000	0.279
9	1.882	1.785	-1.768	0.244	-1.720	0.599	-0.270	0.534	0.200	0.110	-0.110	0.000	0.243
10	1.370	0.664	-0.270	0.607	-2.568	1.486	-0.358	1.442	0.200	0.030	0.030	0.000	0.224
11	1.620	0.632	-0.230	0.589	-1.307	0.957	0.103	0.951	0.200	0.030	0.030	0.000	0.247
12	1.444	0.801	-0.266	0.756	0.800	8.491	2.930	7.970	0.200	0.090	-0.090	0.000	0.524
13	1.755	0.509	0.245	0.447	-1.754	0.594	0.226	0.550	0.200	0.010	0.010	0.000	0.213
14	1.438	0.755	-0.342	0.674	-0.561	4.043	0.439	4.019	0.200	0.110	-0.110	0.000	0.432
15	0.944	0.688	0.174	0.666	1.531	7.422	3.651	6.462	0.200	0.050	0.050	0.000	0.679
16	1.242	0.504	0.062	0.500	-0.545	0.808	0.385	0.710	0.200	0.050	0.050	0.000	0.269
17	1.747	0.625	-0.233	0.580	-0.943	4.219	1.498	3.944	0.200	0.020	-0.020	0.000	0.511
18	1.743	0.507	-0.057	0.504	-1.407	1.070	0.133	1.062	0.200	0.030	0.030	0.000	0.236
19	1.058	0.679	-0.212	0.645	-1.193	1.342	-0.613	1.193	0.200	0.110	-0.110	0.000	0.294
20	0.573	0.519	0.313	0.415	0.403	7.194	4.223	5.824	0.200	0.000	0.000	0.000	0.537
21	1.445	0.563	0.165	0.538	-0.104	0.690	0.336	0.603	0.200	0.080	0.080	0.000	0.288
22	1.493	0.654	-0.198	0.623	-1.248	1.148	-0.018	1.148	0.200	0.080	0.080	0.000	0.338
23	0.952	0.654	-0.438	0.486	0.853	1.449	0.593	1.322	0.200	0.040	-0.040	0.000	0.450
24	1.258	0.835	0.288	0.784	-0.439	5.044	1.431	4.837	0.200	0.040	0.040	0.000	0.506
25	1.768	0.706	-0.453	0.543	0.767	8.147	2.437	7.774	0.200	0.010	0.010	0.000	0.442
26	1.062	0.674	0.072	0.671	-0.246	3.674	0.634	3.619	0.200	0.020	-0.020	0.000	0.383
27	1.486	0.676	-0.174	0.653	0.067	4.210	1.147	4.051	0.200	0.040	0.040	0.000	0.346
28	0.775	0.499	-0.125	0.483	1.670	1.561	1.210	0.986	0.200	0.060	0.060	0.000	0.332
29	1.740	0.970	-0.820	0.519	-0.549	3.856	1.411	3.589	0.200	0.060	0.060	0.000	0.487
30	1.553	0.451	0.013	0.451	-0.890	0.549	0.180	0.518	0.200	0.110	0.110	0.000	0.245
31	1.428	0.792	-0.502	0.612	-0.308	0.703	0.413	0.569	0.200	0.140	0.140	0.000	0.334

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.358	0.929	-0.613	0.698	-0.378	0.643	-0.148	0.626	0.200	0.070	-0.070	0.000	0.320
33	1.120	0.764	-0.560	0.519	-0.238	0.757	0.302	0.695	0.200	0.130	0.130	0.000	0.406
34	1.868	0.863	-0.743	0.440	-1.884	0.760	-0.134	0.749	0.200	0.030	0.030	0.000	0.193
35	1.828	0.820	-0.632	0.523	0.243	6.470	2.053	6.135	0.200	0.050	0.050	0.000	0.277
36	1.060	1.286	-1.070	0.714	3.470	3.366	2.030	2.685	0.200	0.040	0.040	0.000	0.485
37	0.643	0.542	0.133	0.526	0.241	1.707	1.021	1.369	0.200	0.010	-0.010	0.000	0.252
38	0.955	0.689	-0.065	0.686	0.179	6.547	1.239	6.428	0.200	0.010	0.010	0.000	0.468
39	0.603	0.652	-0.358	0.546	3.063	4.409	1.463	4.160	0.200	0.030	-0.030	0.000	0.499
40	1.271	0.699	-0.079	0.695	-0.512	0.672	-0.232	0.631	0.200	0.080	-0.080	0.000	0.263
41	1.264	0.772	0.414	0.651	-0.378	4.265	1.573	3.965	0.200	0.020	0.020	0.000	0.533
42	1.723	0.582	-0.318	0.488	-1.390	0.840	-0.150	0.826	0.200	0.030	-0.030	0.000	0.318
43	0.729	0.629	-0.421	0.467	0.756	0.768	0.136	0.756	0.200	0.140	-0.140	0.000	0.297
44	0.715	0.640	0.115	0.630	-1.601	4.768	0.429	4.748	0.200	0.000	0.000	0.000	0.261
45	1.891	0.979	-0.909	0.362	-1.937	0.768	-0.507	0.577	0.200	0.090	-0.090	0.000	0.211
46	1.499	0.716	-0.201	0.687	0.268	0.468	0.348	0.313	0.200	0.120	0.120	0.000	0.436
47	0.954	0.760	0.154	0.744	5.600	7.214	3.760	6.157	0.200	0.080	0.080	0.000	0.323
48	0.909	0.597	0.009	0.597	-0.745	2.356	-0.175	2.349	0.200	0.090	0.090	0.000	0.292
49	1.182	0.660	0.392	0.531	0.490	1.246	0.820	0.937	0.200	0.070	0.070	0.000	0.317
50	0.473	0.448	0.163	0.418	2.975	3.561	1.535	3.213	0.200	0.010	-0.010	0.000	0.307

Table E.41. Item-Level Evaluation Criteria Summary Statistics, Condition 4, Sample Size = 25

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	1.654	0.573	0.414	0.397	-1.030	0.819	0.720	0.390	0.200	0.070	0.070	0.000	0.267
2	0.790	0.613	-0.300	0.535	-0.819	1.260	-0.529	1.143	0.200	0.130	-0.130	0.000	0.301
3	1.524	0.586	0.154	0.565	-1.459	0.973	0.421	0.877	0.200	0.060	0.060	0.000	0.286
4	1.636	0.690	0.366	0.584	-1.526	1.964	0.424	1.917	0.200	0.050	0.050	0.000	0.349
5	1.151	0.797	0.361	0.710	-1.911	1.920	1.029	1.621	0.200	0.020	0.020	0.000	0.494
6	1.368	0.718	0.108	0.710	-2.064	1.055	0.467	0.947	0.200	0.020	0.020	0.000	0.352
7	1.718	0.507	0.048	0.505	-1.060	0.515	0.320	0.403	0.200	0.060	0.060	0.000	0.237
8	1.911	0.842	-0.809	0.232	-0.989	0.420	0.261	0.329	0.200	0.070	0.070	0.000	0.219
9	1.882	1.784	-1.768	0.243	-1.247	0.350	0.203	0.285	0.200	0.110	-0.110	0.000	0.275
10	1.369	0.664	-0.271	0.606	-1.766	0.935	0.444	0.822	0.200	0.030	0.030	0.000	0.303
11	1.620	0.632	-0.230	0.588	-0.967	0.751	0.443	0.607	0.200	0.030	0.030	0.000	0.274
12	1.444	0.801	-0.266	0.756	-3.283	3.616	-1.153	3.427	0.200	0.090	-0.090	0.000	0.318
13	1.755	0.509	0.245	0.446	-1.283	0.779	0.697	0.349	0.200	0.010	0.010	0.000	0.320
14	1.439	0.755	-0.341	0.673	-1.685	1.985	-0.685	1.863	0.200	0.110	-0.110	0.000	0.282
15	0.944	0.688	0.174	0.665	-1.243	1.487	0.877	1.201	0.200	0.050	0.050	0.000	0.441
16	1.241	0.503	0.061	0.499	-0.486	0.608	0.445	0.415	0.200	0.050	0.050	0.000	0.199
17	1.746	0.626	-0.234	0.581	-1.982	1.708	0.458	1.646	0.200	0.020	-0.020	0.000	0.404
18	1.744	0.506	-0.056	0.503	-1.037	0.687	0.503	0.469	0.200	0.030	0.030	0.000	0.292
19	1.060	0.677	-0.210	0.644	-0.861	0.773	-0.281	0.720	0.200	0.110	-0.110	0.000	0.292
20	0.572	0.520	0.312	0.416	-0.893	3.100	2.927	1.022	0.200	0.000	0.000	0.000	0.432
21	1.446	0.563	0.166	0.538	-0.166	0.431	0.274	0.333	0.200	0.080	0.080	0.000	0.232
22	1.493	0.654	-0.197	0.623	-0.929	0.677	0.301	0.606	0.200	0.080	0.080	0.000	0.288
23	0.952	0.654	-0.438	0.486	0.506	0.863	0.246	0.827	0.200	0.040	-0.040	0.000	0.347
24	1.258	0.835	0.288	0.784	-1.865	2.283	0.005	2.283	0.200	0.040	0.040	0.000	0.391
25	1.767	0.706	-0.453	0.542	-1.197	0.563	0.473	0.306	0.200	0.010	0.010	0.000	0.312
26	1.062	0.674	0.072	0.671	-1.469	1.689	-0.589	1.583	0.200	0.020	-0.020	0.000	0.246
27	1.486	0.676	-0.174	0.653	-1.437	2.024	-0.357	1.992	0.200	0.040	0.040	0.000	0.240
28	0.775	0.499	-0.125	0.483	1.052	0.835	0.592	0.589	0.200	0.060	0.060	0.000	0.258
29	1.740	0.970	-0.820	0.519	-1.645	1.588	0.315	1.556	0.200	0.060	0.060	0.000	0.404
30	1.554	0.451	0.014	0.451	-0.720	0.485	0.350	0.336	0.200	0.110	0.110	0.000	0.204
31	1.428	0.792	-0.502	0.612	-0.296	0.652	0.424	0.495	0.200	0.140	0.140	0.000	0.318

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.356	0.931	-0.614	0.699	-0.335	0.325	-0.105	0.307	0.200	0.070	-0.070	0.000	0.278
33	1.120	0.763	-0.560	0.518	-0.268	0.536	0.272	0.462	0.200	0.130	0.130	0.000	0.369
34	1.868	0.862	-0.742	0.439	-1.305	0.653	0.445	0.478	0.200	0.030	0.030	0.000	0.286
35	1.828	0.820	-0.632	0.523	-1.162	0.721	0.648	0.316	0.200	0.050	0.050	0.000	0.297
36	1.060	1.286	-1.070	0.714	2.478	2.279	1.038	2.030	0.200	0.040	0.040	0.000	0.459
37	0.644	0.541	0.134	0.524	-0.085	0.849	0.695	0.489	0.200	0.010	-0.010	0.000	0.236
38	0.955	0.689	-0.065	0.686	-1.285	0.632	-0.225	0.591	0.200	0.010	0.010	0.000	0.302
39	0.603	0.652	-0.357	0.546	0.886	2.358	-0.714	2.247	0.200	0.030	-0.030	0.000	0.517
40	1.271	0.698	-0.079	0.694	-0.342	0.317	-0.062	0.311	0.200	0.080	-0.080	0.000	0.243
41	1.265	0.772	0.415	0.652	-1.647	1.917	0.303	1.893	0.200	0.020	0.020	0.000	0.357
42	1.723	0.581	-0.317	0.487	-1.029	0.516	0.211	0.471	0.200	0.030	-0.030	0.000	0.262
43	0.730	0.628	-0.420	0.467	0.442	0.484	-0.178	0.450	0.200	0.140	-0.140	0.000	0.234
44	0.716	0.640	0.116	0.629	-1.840	1.167	0.190	1.152	0.200	0.000	0.000	0.000	0.253
45	1.891	0.979	-0.909	0.362	-1.352	0.367	0.079	0.359	0.200	0.090	-0.090	0.000	0.216
46	1.500	0.716	-0.200	0.687	0.147	0.415	0.227	0.348	0.200	0.120	0.120	0.000	0.445
47	0.956	0.760	0.156	0.743	1.787	2.066	-0.053	2.066	0.200	0.080	0.080	0.000	0.393
48	0.910	0.596	0.010	0.596	-0.151	0.746	0.419	0.618	0.200	0.090	0.090	0.000	0.258
49	1.181	0.660	0.391	0.531	0.245	0.776	0.575	0.520	0.200	0.070	0.070	0.000	0.230
50	0.473	0.449	0.163	0.419	0.884	0.934	-0.556	0.750	0.200	0.010	-0.010	0.000	0.278

Table E.42. Item-Level Evaluation Criteria Summary Statistics, Condition 5, Sample Size = 25

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
1	0.936	0.384	-0.304	0.235	-1.764	0.966	-0.014	0.966	0.197	0.067	0.067	0.005	0.259
2	0.752	0.776	-0.338	0.699	-0.655	1.378	-0.365	1.329	0.195	0.136	-0.135	0.011	0.370
3	0.790	0.657	-0.580	0.309	-2.587	1.283	-0.707	1.070	0.198	0.059	0.058	0.004	0.291
4	0.965	0.514	-0.305	0.414	-2.283	2.110	-0.333	2.083	0.195	0.046	0.045	0.008	0.257
5	0.664	0.431	-0.126	0.412	-3.653	2.296	-0.713	2.182	0.199	0.019	0.019	0.003	0.222
6	0.743	0.581	-0.517	0.267	-3.478	1.518	-0.948	1.186	0.200	0.020	0.020	0.000	0.166
7	1.018	0.805	-0.653	0.472	-1.811	0.941	-0.431	0.837	0.194	0.055	0.054	0.008	0.372
8	1.384	1.385	-1.336	0.364	-1.482	0.719	-0.232	0.681	0.191	0.061	0.061	0.008	0.322
9	1.150	2.520	-2.500	0.320	-2.174	1.071	-0.724	0.789	0.197	0.113	-0.113	0.005	0.285
10	0.744	0.949	-0.896	0.313	-3.487	2.396	-1.277	2.028	0.199	0.029	0.029	0.003	0.264
11	1.073	0.874	-0.777	0.400	-1.449	1.234	-0.039	1.234	0.195	0.026	0.025	0.008	0.312
12	0.797	1.008	-0.913	0.425	-4.101	2.957	-1.971	2.205	0.197	0.094	-0.093	0.006	0.306
13	0.880	0.676	-0.630	0.245	-2.449	0.864	-0.469	0.725	0.198	0.009	0.008	0.004	0.215
14	0.968	0.890	-0.813	0.363	-2.097	2.356	-1.097	2.085	0.196	0.114	-0.114	0.008	0.268
15	0.571	0.356	-0.199	0.295	-3.143	2.714	-1.023	2.514	0.198	0.048	0.048	0.004	0.313
16	0.778	0.519	-0.402	0.329	-0.652	0.883	0.278	0.838	0.192	0.042	0.042	0.008	0.322
17	0.914	1.102	-1.066	0.278	-3.155	1.419	-0.715	1.226	0.200	0.020	-0.020	0.000	0.218
18	1.165	0.743	-0.635	0.385	-1.468	0.537	0.073	0.532	0.194	0.026	0.024	0.009	0.232
19	0.664	0.644	-0.606	0.219	-0.977	0.986	-0.397	0.903	0.197	0.114	-0.113	0.008	0.273
20	0.448	0.291	0.188	0.221	-2.001	2.589	1.819	1.842	0.199	0.005	-0.001	0.005	0.328
21	1.085	0.553	-0.195	0.517	0.088	0.766	0.528	0.555	0.186	0.067	0.066	0.010	0.292
22	1.039	0.820	-0.651	0.499	-1.358	1.098	-0.128	1.091	0.189	0.070	0.069	0.011	0.366
23	0.684	0.747	-0.706	0.246	1.259	1.710	0.999	1.388	0.188	0.055	-0.053	0.015	0.510
24	0.886	0.565	-0.084	0.558	-2.104	2.091	-0.234	2.078	0.193	0.034	0.033	0.008	0.339
25	1.139	1.149	-1.081	0.389	-2.431	1.882	-0.761	1.721	0.198	0.009	0.008	0.004	0.356
26	0.667	0.395	-0.323	0.227	-1.500	1.120	-0.620	0.933	0.197	0.024	-0.023	0.006	0.282
27	1.118	0.743	-0.543	0.508	-1.718	1.862	-0.638	1.749	0.193	0.035	0.033	0.011	0.331
28	0.569	0.369	-0.331	0.164	2.178	2.209	1.718	1.389	0.184	0.046	0.044	0.011	0.369
29	1.142	1.493	-1.418	0.465	-2.540	1.813	-0.580	1.718	0.195	0.055	0.055	0.007	0.356
30	1.116	0.691	-0.424	0.545	-1.114	0.923	-0.044	0.922	0.190	0.101	0.100	0.011	0.353
31	1.155	0.973	-0.775	0.589	-0.232	1.040	0.488	0.918	0.186	0.126	0.126	0.011	0.404

Item ID	a_Average	a_RMSE	a_Bias	a_SE	b_Average	b_RMSE	b_Bias	b_SE	c_Average	c_RMSE	c_Bias	c_SE	ICC_RMSE
32	1.037	1.093	-0.933	0.570	-0.163	0.556	0.068	0.552	0.191	0.080	-0.079	0.012	0.347
33	0.833	0.905	-0.847	0.321	-0.144	0.820	0.396	0.718	0.192	0.122	0.122	0.011	0.450
34	1.205	1.451	-1.405	0.363	-2.154	0.949	-0.404	0.859	0.195	0.027	0.025	0.009	0.318
35	1.056	1.470	-1.404	0.434	-2.518	1.572	-0.708	1.404	0.195	0.046	0.045	0.008	0.383
36	0.514	1.634	-1.616	0.241	4.308	3.633	2.868	2.230	0.178	0.025	0.018	0.017	0.511
37	0.469	0.239	-0.041	0.236	0.298	1.404	1.078	0.899	0.195	0.017	-0.015	0.008	0.270
38	0.648	0.500	-0.373	0.333	-2.364	2.291	-1.304	1.883	0.198	0.010	0.008	0.006	0.352
39	0.447	0.565	-0.513	0.235	3.718	3.288	2.118	2.515	0.191	0.042	-0.039	0.014	0.345
40	1.008	0.784	-0.343	0.705	-0.284	0.553	-0.004	0.553	0.187	0.094	-0.093	0.013	0.261
41	0.869	0.373	0.019	0.373	-1.842	1.316	0.108	1.311	0.198	0.020	0.018	0.007	0.301
42	1.154	0.967	-0.886	0.388	-1.680	1.131	-0.440	1.042	0.193	0.038	-0.037	0.009	0.357
43	0.654	0.711	-0.496	0.509	1.079	0.762	0.459	0.608	0.193	0.147	-0.147	0.012	0.333
44	0.514	0.325	-0.086	0.314	-3.177	2.459	-1.147	2.175	0.199	0.003	-0.001	0.003	0.218
45	1.158	1.671	-1.643	0.309	-2.352	1.217	-0.922	0.795	0.197	0.094	-0.093	0.005	0.259
46	1.271	0.741	-0.429	0.604	0.584	0.894	0.664	0.599	0.178	0.099	0.098	0.015	0.451
47	0.503	0.425	-0.298	0.303	4.523	3.630	2.683	2.444	0.183	0.066	0.063	0.019	0.314
48	0.668	0.412	-0.232	0.341	-0.104	1.241	0.466	1.150	0.191	0.081	0.081	0.009	0.317
49	0.813	0.383	0.023	0.382	0.854	1.555	1.184	1.007	0.183	0.054	0.053	0.011	0.346
50	0.433	0.251	0.123	0.219	2.176	1.501	0.736	1.308	0.194	0.019	-0.016	0.011	0.276

Appendix F

Average Test Characteristic Curves

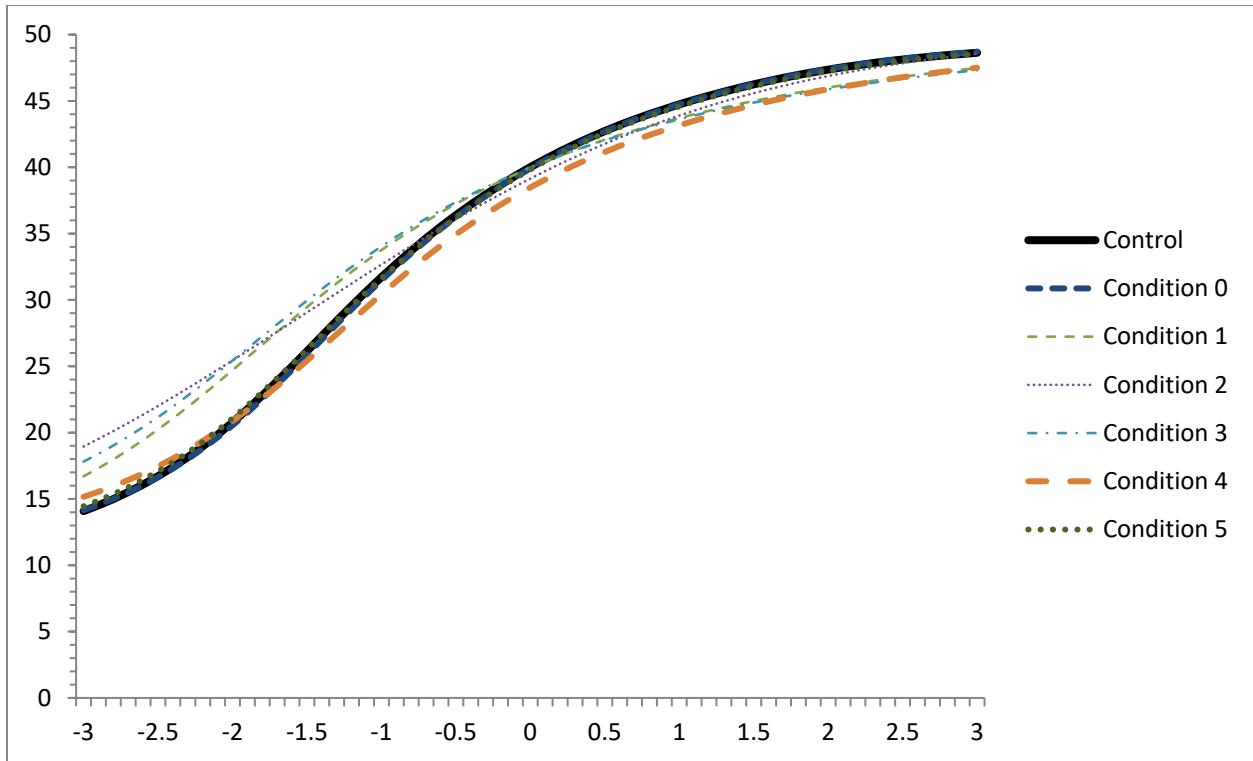


Figure F.1. Test Characteristic Curves by Condition, Sample Size = 1,000

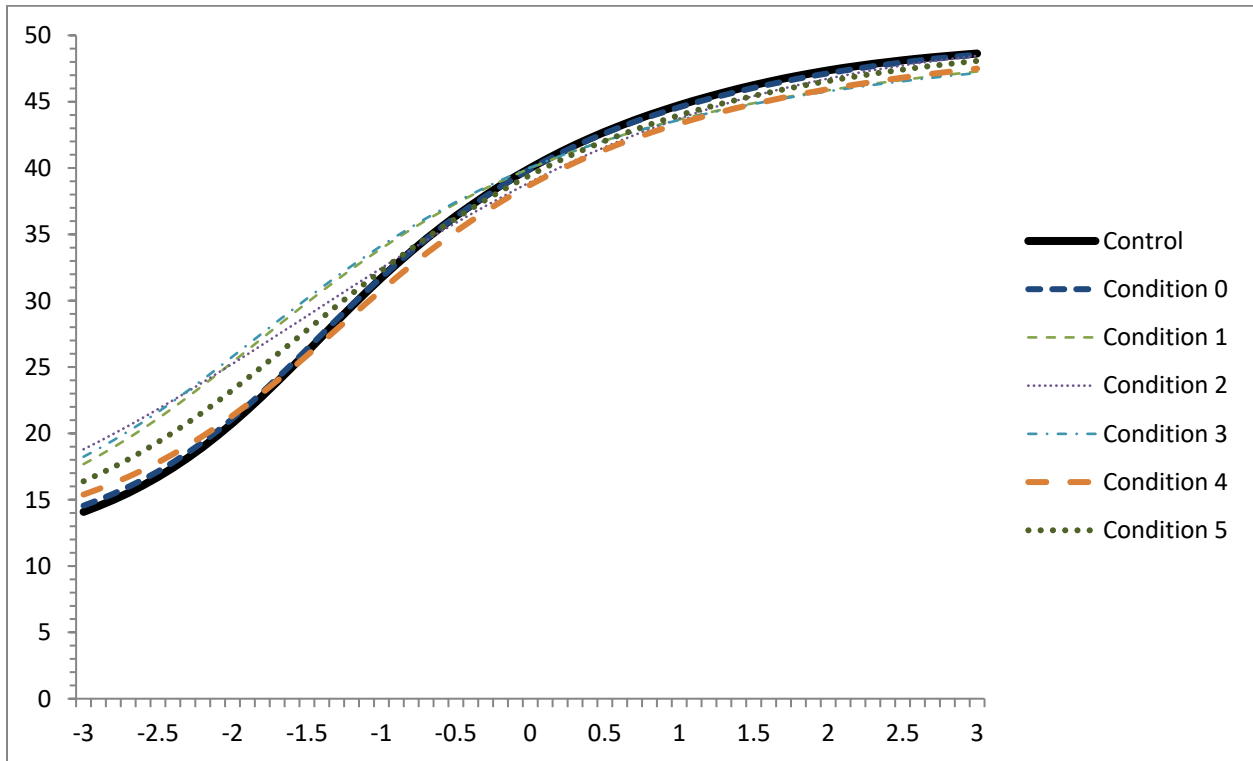


Figure F.2. Test Characteristic Curves by Condition, Sample Size = 200

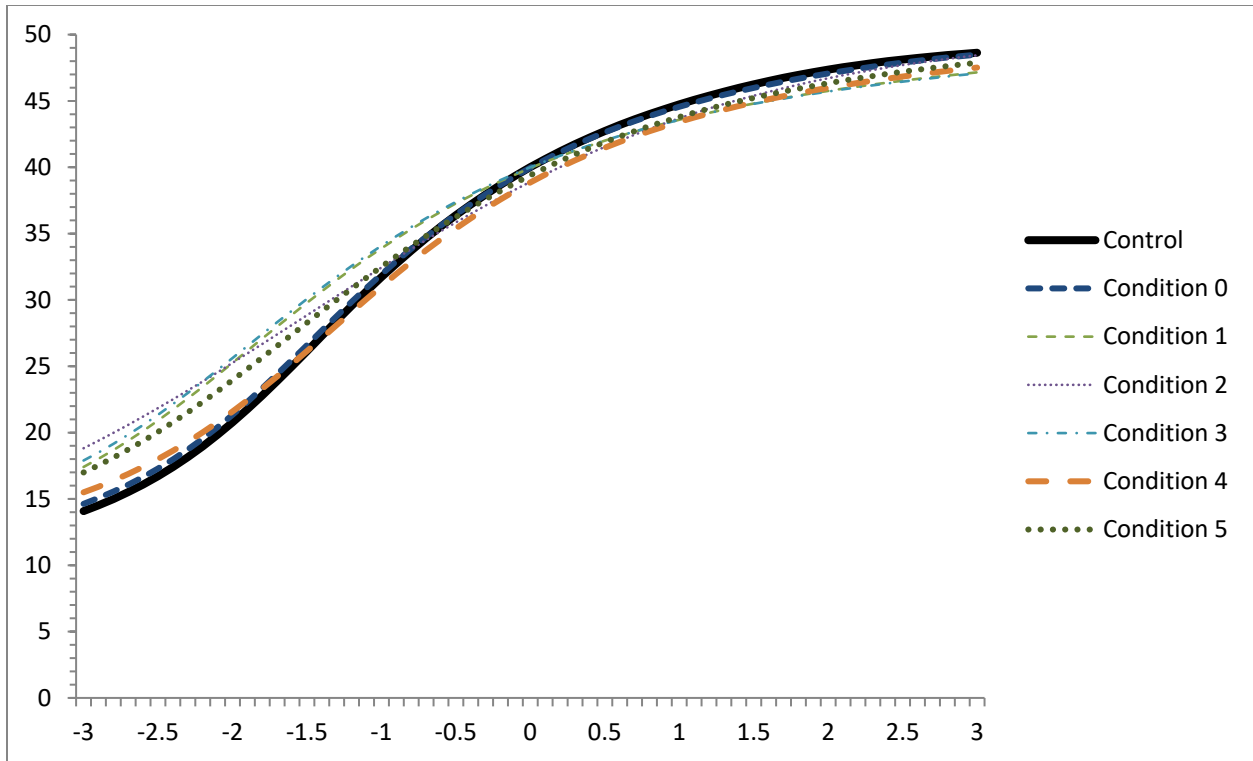


Figure F.3. Test Characteristic Curves by Condition, Sample Size = 150

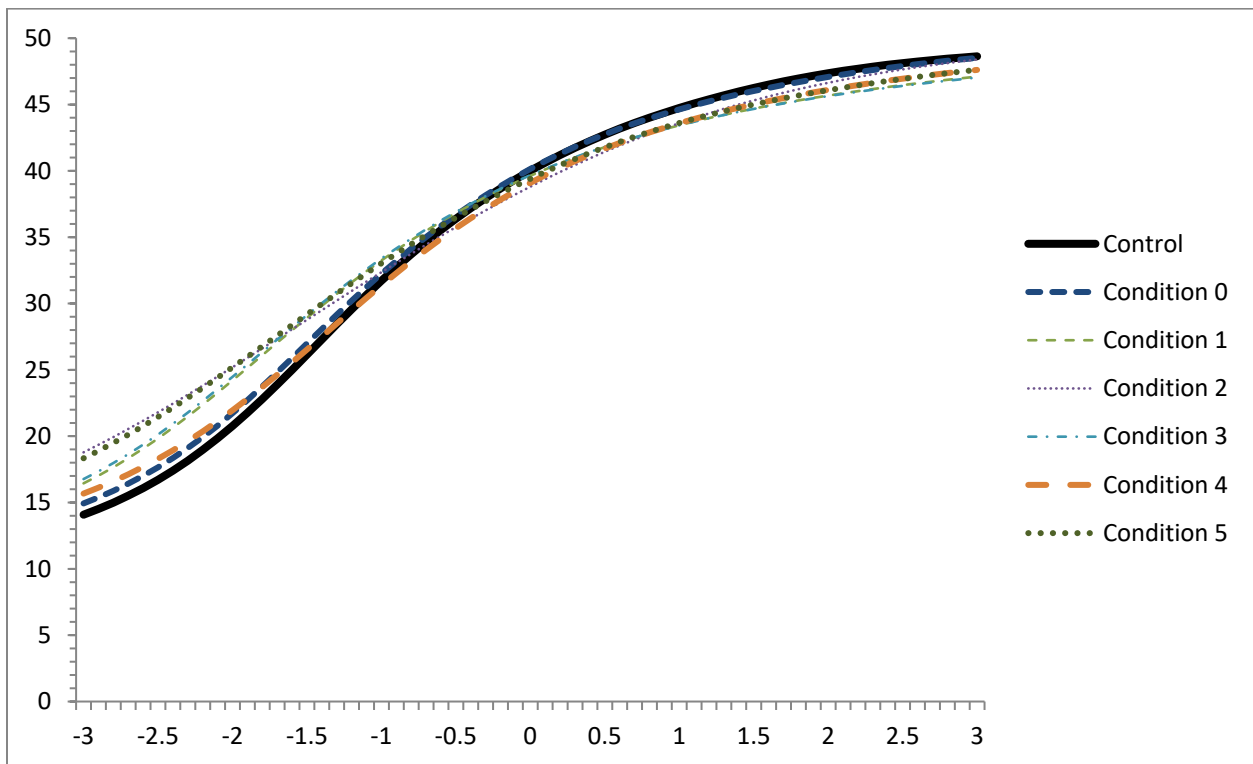


Figure F.4. Test Characteristic Curves by Condition, Sample Size = 100

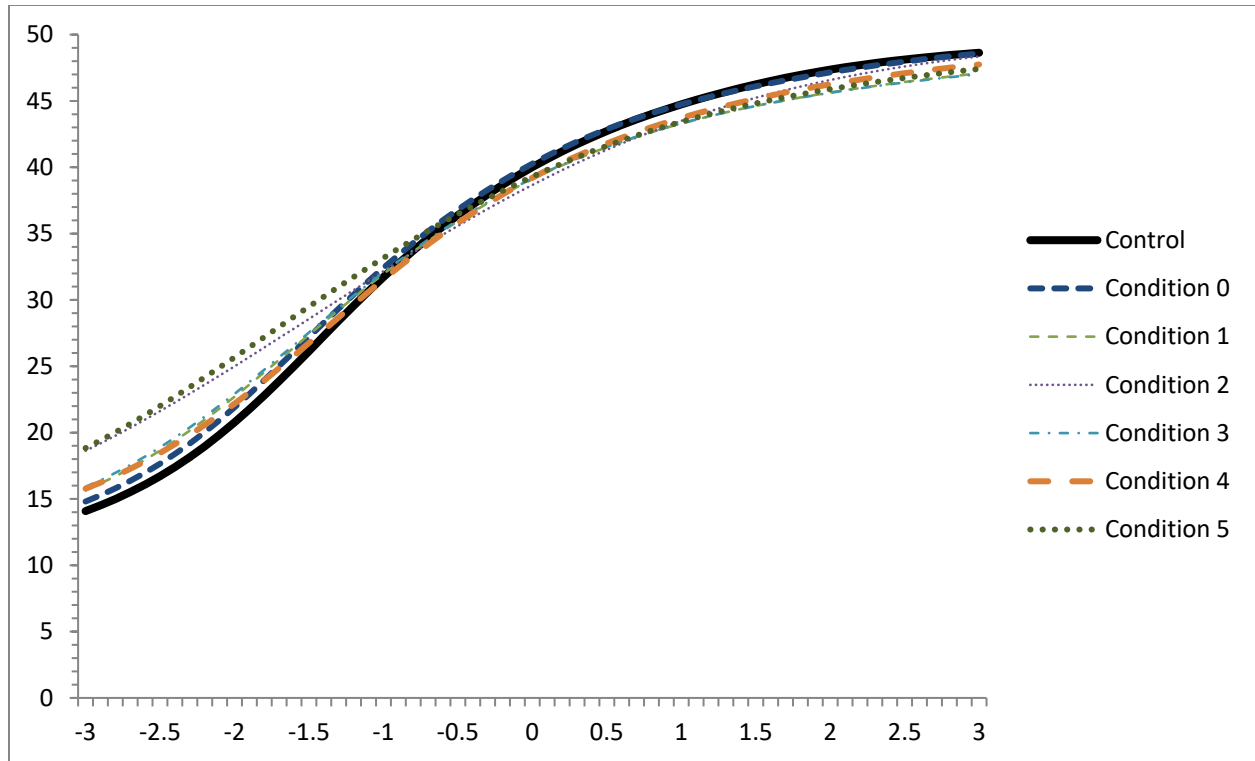


Figure F.5. Test Characteristic Curves by Condition, Sample Size = 75

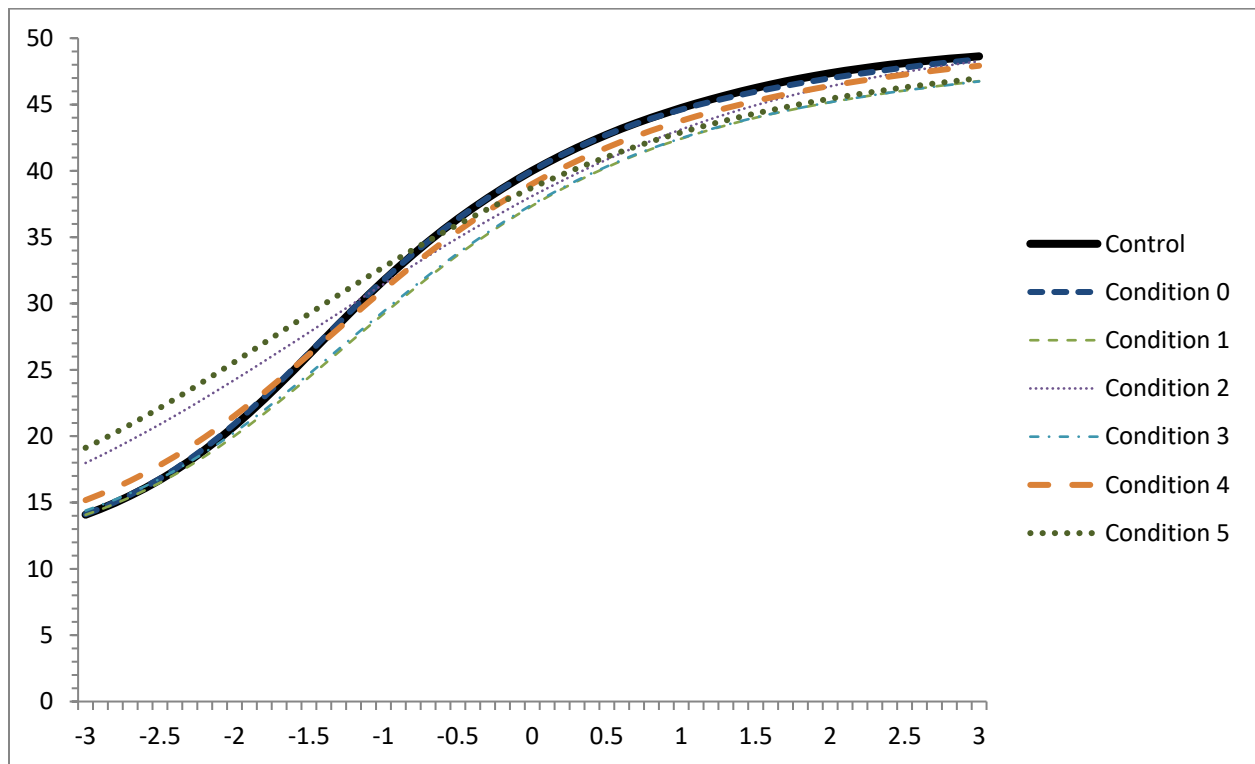


Figure F.6. Test Characteristic Curves by Condition, Sample Size = 50

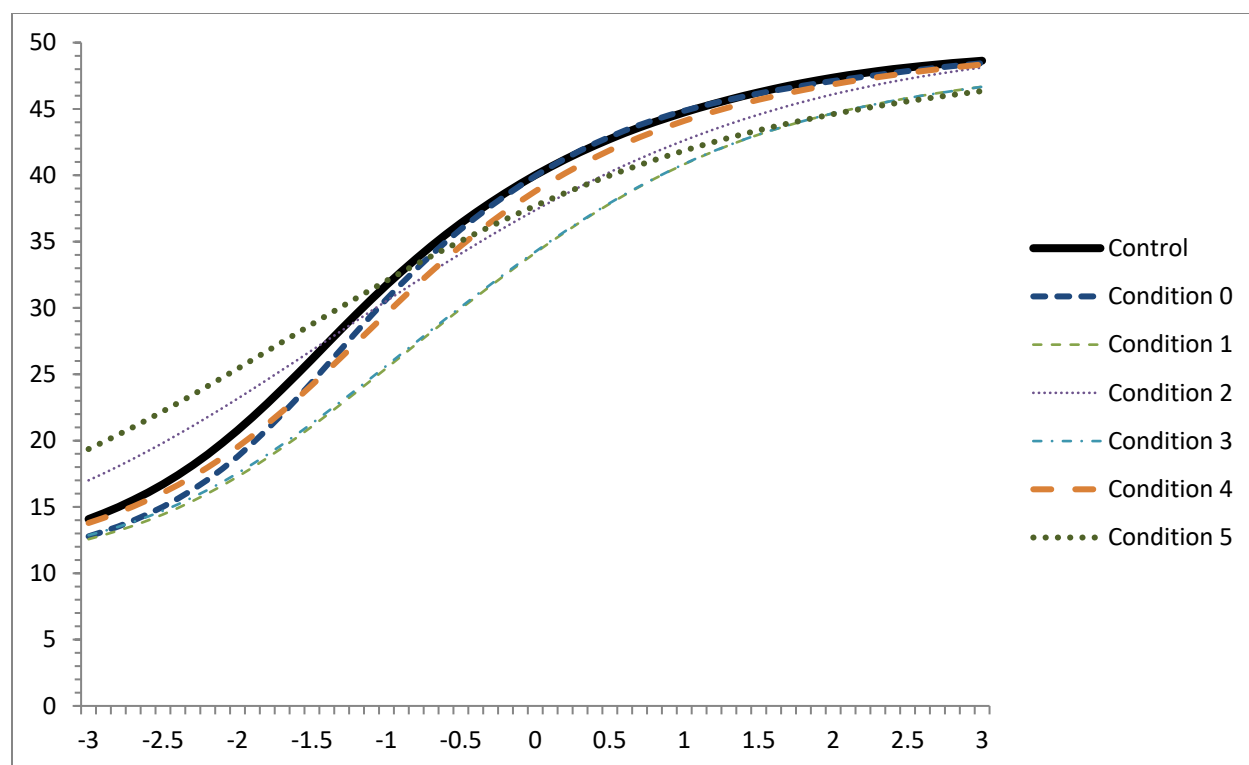


Figure F.7. Test Characteristic Curves by Condition, Sample Size = 25

Appendix G

Classification Accuracy and Reliability

Table G.1. Decision Accuracy and Reliability, by Condition, Sample Size = 1,000

ILR Level	Control Probability	Study Average Probability Condition 0	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 1	Study - Control
0	0.002	0.002	0.000	0	0.002	0.000	-0.002
0+	0.003	0.003	0.000	0+	0.003	0.001	-0.002
1	0.040	0.043	0.002	1	0.040	0.020	-0.020
1+	0.096	0.099	0.004	1+	0.096	0.076	-0.020
2	0.227	0.229	0.002	2	0.227	0.249	0.022
2+	0.227	0.225	-0.002	2+	0.227	0.282	0.055
3	0.405	0.399	-0.007	3	0.405	0.372	-0.033
Reliability	0.848	0.851	0.003	Reliability	0.848	0.777	-0.072

ILR Level	Control Probability	Study Average Probability Condition 2	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 3	Study - Control
0	0.002	0.000	-0.002	0	0.002	0.000	-0.002
0+	0.003	0.001	-0.003	0+	0.003	0.001	-0.003
1	0.040	0.017	-0.023	1	0.040	0.017	-0.023
1+	0.096	0.073	-0.023	1+	0.096	0.073	-0.023
2	0.227	0.250	0.023	2	0.227	0.250	0.023
2+	0.227	0.289	0.062	2+	0.227	0.289	0.062
3	0.405	0.371	-0.035	3	0.405	0.371	-0.035
Reliability	0.848	0.765	-0.084	Reliability	0.848	0.765	-0.084

ILR Level	Control Probability	Study Average Probability Condition 4	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 5	Study - Control
0	0.002	0.002	0.000	0	0.002	0.002	0.000
0+	0.003	0.003	0.000	0+	0.003	0.003	0.000
1	0.040	0.048	0.008	1	0.040	0.040	0.000
1+	0.096	0.122	0.026	1+	0.096	0.100	0.004
2	0.227	0.276	0.050	2	0.227	0.235	0.008
2+	0.227	0.250	0.024	2+	0.227	0.230	0.003
3	0.405	0.298	-0.108	3	0.405	0.391	-0.014
Reliability	0.848	0.808	-0.041	Reliability	0.848	0.845	-0.003

Table G.2. Decision Accuracy and Reliability, by Condition, Sample Size = 200

ILR Level	Control Probability	Study Average Probability Condition 0	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 1	Study - Control
0	0.002	0.002	0.000	0	0.002	0.000	-0.002
0+	0.003	0.003	0.000	0+	0.003	0.001	-0.003
1	0.040	0.043	0.003	1	0.040	0.021	-0.019
1+	0.096	0.104	0.009	1+	0.096	0.082	-0.014
2	0.227	0.237	0.010	2	0.227	0.258	0.031
2+	0.227	0.229	0.002	2+	0.227	0.284	0.057
3	0.405	0.381	-0.024	3	0.405	0.355	-0.051
Reliability	0.848	0.845	-0.003	Reliability	0.848	0.774	-0.075
ILR Level	Control Probability	Study Average Probability Condition 2	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 3	Study - Control
0	0.002	0.000	-0.002	0	0.002	0.000	-0.002
0+	0.003	0.000	-0.003	0+	0.003	0.001	-0.003
1	0.040	0.019	-0.021	1	0.040	0.019	-0.021
1+	0.096	0.101	0.005	1+	0.096	0.079	-0.017
2	0.227	0.292	0.065	2	0.227	0.258	0.031
2+	0.227	0.256	0.030	2+	0.227	0.288	0.062
3	0.405	0.332	-0.073	3	0.405	0.355	-0.050
Reliability	0.848	0.812	-0.036	Reliability	0.848	0.765	-0.083
ILR Level	Control Probability	Study Average Probability Condition 4	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 5	Study - Control
0	0.002	0.002	0.000	0	0.002	0.001	-0.002
0+	0.003	0.003	0.000	0+	0.003	0.002	-0.002
1	0.040	0.048	0.008	1	0.040	0.033	-0.008
1+	0.096	0.123	0.028	1+	0.096	0.105	0.010
2	0.227	0.276	0.049	2	0.227	0.265	0.038
2+	0.227	0.250	0.023	2+	0.227	0.249	0.022
3	0.405	0.298	-0.108	3	0.405	0.347	-0.059
Reliability	0.848	0.804	-0.044	Reliability	0.848	0.823	-0.025

Table G.3. Decision Accuracy and Reliability, by Condition, Sample Size = 150

ILR Level	Control Probability	Study Average Probability Condition 0	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 1	Study - Control
0	0.002	0.002	0.000	0	0.002	0.000	-0.002
0+	0.003	0.003	0.000	0+	0.003	0.001	-0.002
1	0.040	0.043	0.003	1	0.040	0.021	-0.019
1+	0.096	0.105	0.010	1+	0.096	0.083	-0.013
2	0.227	0.239	0.012	2	0.227	0.260	0.034
2+	0.227	0.230	0.004	2+	0.227	0.284	0.057
3	0.405	0.378	-0.028	3	0.405	0.351	-0.055
Reliability	0.848	0.843	-0.005	Reliability	0.848	0.773	-0.076
ILR Level	Control Probability	Study Average Probability Condition 2	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 3	Study - Control
0	0.002	0.000	-0.002	0	0.002	0.000	-0.002
0+	0.003	0.000	-0.003	0+	0.003	0.001	-0.003
1	0.040	0.019	-0.021	1	0.040	0.019	-0.021
1+	0.096	0.102	0.006	1+	0.096	0.080	-0.015
2	0.227	0.294	0.067	2	0.227	0.260	0.033
2+	0.227	0.256	0.030	2+	0.227	0.288	0.061
3	0.405	0.329	-0.077	3	0.405	0.352	-0.054
Reliability	0.848	0.812	-0.037	Reliability	0.848	0.765	-0.083
ILR Level	Control Probability	Study Average Probability Condition 4	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 5	Study - Control
0	0.002	0.002	0.000	0	0.002	0.001	-0.002
0+	0.003	0.003	0.000	0+	0.003	0.001	-0.002
1	0.040	0.048	0.008	1	0.040	0.030	-0.011
1+	0.096	0.123	0.028	1+	0.096	0.106	0.010
2	0.227	0.276	0.049	2	0.227	0.274	0.047
2+	0.227	0.250	0.024	2+	0.227	0.255	0.028
3	0.405	0.298	-0.107	3	0.405	0.335	-0.071
Reliability	0.848	0.802	-0.046	Reliability	0.848	0.814	-0.034

Table G.4. Decision Accuracy and Reliability, by Condition, Sample Size = 100

ILR Level	Control Probability	Study Average Probability Condition 0	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 1	Study - Control
0	0.002	0.002	0.000	0	0.002	0.000	-0.002
0+	0.003	0.003	0.000	0+	0.003	0.001	-0.002
1	0.040	0.046	0.006	1	0.040	0.024	-0.017
1+	0.096	0.111	0.016	1+	0.096	0.090	-0.006
2	0.227	0.244	0.018	2	0.227	0.268	0.042
2+	0.227	0.232	0.006	2+	0.227	0.285	0.058
3	0.405	0.372	-0.033	3	0.405	0.343	-0.062
Reliability	0.848	0.852	0.004	Reliability	0.848	0.784	-0.064
ILR Level	Control Probability	Study Average Probability Condition 2	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 3	Study - Control
0	0.002	0.000	-0.002	0	0.002	0.000	-0.002
0+	0.003	0.001	-0.003	0+	0.003	0.001	-0.003
1	0.040	0.021	-0.020	1	0.040	0.022	-0.019
1+	0.096	0.108	0.012	1+	0.096	0.087	-0.009
2	0.227	0.302	0.075	2	0.227	0.268	0.041
2+	0.227	0.258	0.032	2+	0.227	0.288	0.061
3	0.405	0.322	-0.084	3	0.405	0.345	-0.060
Reliability	0.848	0.821	-0.028	Reliability	0.848	0.777	-0.071
ILR Level	Control Probability	Study Average Probability Condition 4	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 5	Study - Control
0	0.002	0.002	0.000	0	0.002	0.000	-0.002
0+	0.003	0.003	0.000	0+	0.003	0.001	-0.002
1	0.040	0.050	0.010	1	0.040	0.027	-0.013
1+	0.096	0.127	0.032	1+	0.096	0.110	0.014
2	0.227	0.279	0.052	2	0.227	0.293	0.066
2+	0.227	0.251	0.025	2+	0.227	0.266	0.039
3	0.405	0.298	-0.108	3	0.405	0.314	-0.091
Reliability	0.848	0.811	-0.037	Reliability	0.848	0.810	-0.038

Table G.5. Decision Accuracy and Reliability, by Condition, Sample Size = 75

ILR Level	Control Probability	Study Average Probability Condition 0	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 1	Study - Control
0	0.002	0.002	0.000	0	0.002	0.001	-0.002
0+	0.003	0.003	0.000	0+	0.003	0.001	-0.002
1	0.040	0.048	0.008	1	0.040	0.026	-0.015
1+	0.096	0.115	0.019	1+	0.096	0.094	-0.001
2	0.227	0.245	0.019	2	0.227	0.271	0.045
2+	0.227	0.231	0.004	2+	0.227	0.281	0.054
3	0.405	0.367	-0.038	3	0.405	0.338	-0.068
Reliability	0.848	0.853	0.005	Reliability	0.848	0.786	-0.062
ILR Level	Control Probability	Study Average Probability Condition 2	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 3	Study - Control
0	0.002	0.000	-0.002	0	0.002	0.000	-0.002
0+	0.003	0.001	-0.003	0+	0.003	0.001	-0.002
1	0.040	0.022	-0.018	1	0.040	0.024	-0.016
1+	0.096	0.112	0.017	1+	0.096	0.092	-0.004
2	0.227	0.306	0.080	2	0.227	0.271	0.044
2+	0.227	0.257	0.030	2+	0.227	0.284	0.057
3	0.405	0.313	-0.092	3	0.405	0.340	-0.066
Reliability	0.848	0.822	-0.027	Reliability	0.848	0.780	-0.068
ILR Level	Control Probability	Study Average Probability Condition 4	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 5	Study - Control
0	0.002	0.002	0.000	0	0.002	0.000	-0.002
0+	0.003	0.003	0.000	0+	0.003	0.001	-0.003
1	0.040	0.052	0.012	1	0.040	0.026	-0.015
1+	0.096	0.130	0.035	1+	0.096	0.112	0.017
2	0.227	0.281	0.054	2	0.227	0.305	0.078
2+	0.227	0.250	0.023	2+	0.227	0.271	0.045
3	0.405	0.293	-0.113	3	0.405	0.296	-0.109
Reliability	0.848	0.811	-0.037	Reliability	0.848	0.800	-0.049

Table G.6. Decision Accuracy and Reliability, by Condition, Sample Size = 50

ILR Level	Control Probability	Study Average Probability Condition 0	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 1	Study - Control
0	0.002	0.003	0.000	0	0.002	0.001	-0.002
0+	0.003	0.004	0.001	0+	0.003	0.002	-0.002
1	0.040	0.058	0.018	1	0.040	0.033	-0.007
1+	0.096	0.127	0.032	1+	0.096	0.111	0.015
2	0.227	0.249	0.022	2	0.227	0.287	0.060
2+	0.227	0.222	-0.005	2+	0.227	0.269	0.042
3	0.405	0.337	-0.068	3	0.405	0.298	-0.107
Reliability	0.848	0.850	0.002	Reliability	0.848	0.792	-0.056
ILR Level	Control Probability	Study Average Probability Condition 2	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 3	Study - Control
0	0.002	0.000	-0.002	0	0.002	0.001	-0.002
0+	0.003	0.001	-0.003	0+	0.003	0.001	-0.002
1	0.040	0.027	-0.013	1	0.040	0.031	-0.009
1+	0.096	0.126	0.030	1+	0.096	0.108	0.013
2	0.227	0.318	0.091	2	0.227	0.287	0.060
2+	0.227	0.249	0.023	2+	0.227	0.272	0.045
3	0.405	0.279	-0.126	3	0.405	0.300	-0.106
Reliability	0.848	0.815	-0.033	Reliability	0.848	0.787	-0.061
ILR Level	Control Probability	Study Average Probability Condition 4	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 5	Study - Control
0	0.002	0.003	0.000	0	0.002	0.000	-0.002
0+	0.003	0.004	0.001	0+	0.003	0.001	-0.003
1	0.040	0.059	0.019	1	0.040	0.027	-0.013
1+	0.096	0.139	0.043	1+	0.096	0.125	0.030
2	0.227	0.281	0.054	2	0.227	0.330	0.104
2+	0.227	0.238	0.012	2+	0.227	0.269	0.042
3	0.405	0.277	-0.129	3	0.405	0.248	-0.158
Reliability	0.848	0.810	-0.038	Reliability	0.848	0.780	-0.068

Table G.7. Decision Accuracy and Reliability, by Condition, Sample Size = 25

ILR Level	Control Probability	Study Average Probability Condition 0	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 1	Study - Control
0	0.002	0.006	0.004	0	0.002	0.002	0.000
0+	0.003	0.008	0.004	0+	0.003	0.003	0.000
1	0.040	0.080	0.039	1	0.040	0.051	0.010
1+	0.096	0.137	0.042	1+	0.096	0.129	0.034
2	0.227	0.235	0.008	2	0.227	0.276	0.050
2+	0.227	0.205	-0.022	2+	0.227	0.245	0.018
3	0.405	0.330	-0.076	3	0.405	0.294	-0.112
Reliability	0.848	0.861	0.013	Reliability	0.848	0.815	-0.033
ILR Level	Control Probability	Study Average Probability Condition 2	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 3	Study - Control
0	0.002	0.001	-0.002	0	0.002	0.002	-0.001
0+	0.003	0.001	-0.002	0+	0.003	0.003	0.000
1	0.040	0.036	-0.005	1	0.040	0.048	0.008
1+	0.096	0.147	0.052	1+	0.096	0.127	0.032
2	0.227	0.336	0.110	2	0.227	0.277	0.050
2+	0.227	0.238	0.011	2+	0.227	0.247	0.021
3	0.405	0.241	-0.164	3	0.405	0.296	-0.110
Reliability	0.848	0.817	-0.031	Reliability	0.848	0.811	-0.037
ILR Level	Control Probability	Study Average Probability Condition 4	Study - Control	ILR Level	Control Probability	Study Average Probability Condition 5	Study - Control
0	0.002	0.005	0.003	0	0.002	0.000	-0.002
0+	0.003	0.007	0.004	0+	0.003	0.001	-0.003
1	0.040	0.077	0.037	1	0.040	0.032	-0.008
1+	0.096	0.149	0.053	1+	0.096	0.155	0.059
2	0.227	0.267	0.040	2	0.227	0.380	0.153
2+	0.227	0.220	-0.007	2+	0.227	0.258	0.032
3	0.405	0.276	-0.130	3	0.405	0.174	-0.231
Reliability	0.848	0.832	-0.017	Reliability	0.848	0.762	-0.087

Appendix H

Classification Consistency and Kappa

Table H.1. Examinee Classification: Control Condition vs. Study Condition 0 Percentages and Kappa, Sample Size = 1,000

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.13	0.00	0.00	0.00	0.00	0.13
	1	0.00	0.05	4.39	0.10	0.00	0.00	4.54
	1+	0.00	0.00	0.12	6.62	0.29	0.00	7.03
	2	0.00	0.00	0.00	0.10	21.22	0.00	21.32
	2+	0.00	0.00	0.00	0.00	0.70	27.27	27.96
	3	0.00	0.00	0.00	0.00	0.00	4.58	34.44
Total	0.00	0.18	4.51	6.82	22.21	31.85	34.44	94.06
<i>kappa</i>								91.78

Table H.2. Examinee Classification: Control Condition vs. Study Condition 1 Percentages and Kappa, Sample Size = 1,000

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.18	0.62	0.00	0.00	0.00	0.79
	1	0.00	0.00	3.90	2.42	0.00	0.00	6.31
	1+	0.00	0.00	0.00	4.40	3.45	0.00	7.85
	2	0.00	0.00	0.00	0.00	18.72	0.30	19.02
	2+	0.00	0.00	0.00	0.00	0.04	24.71	24.75
	3	0.00	0.00	0.00	0.00	0.00	6.83	34.44
Total	0.00	0.18	4.51	6.82	22.21	31.85	34.44	86.34
<i>kappa</i>								81.24

Table H.3. Examinee Classification: Control Condition vs. Study Condition 2 Percentages and Kappa, Sample Size = 1,000

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.18	0.92	0.00	0.00	0.00	1.10
	1	0.00	0.00	3.60	2.27	0.00	0.00	5.86
	1+	0.00	0.00	0.00	4.55	2.04	0.00	6.59
	2	0.00	0.00	0.00	0.00	18.03	0.00	18.03
	2+	0.00	0.00	0.00	0.00	2.14	24.60	26.74
	3	0.00	0.00	0.00	0.00	0.00	7.24	34.44
Total	0.00	0.18	4.51	6.82	22.21	31.85	34.44	85.39
<i>kappa</i>								79.81

Table H.4. Examinee Classification: Control Condition vs. Study Condition 3 Percentages and Kappa, Sample Size = 1,000

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.18	0.85	0.00	0.00	0.00	1.03
	1	0.00	0.00	3.66	2.94	0.00	0.00	6.60
	1+	0.00	0.00	0.00	3.88	4.00	0.00	7.88
	2	0.00	0.00	0.00	0.00	18.21	0.51	18.72
	2+	0.00	0.00	0.00	0.00	0.00	24.51	24.51
	3	0.00	0.00	0.00	0.00	0.00	6.83	34.44
Total	0.00	0.18	4.51	6.82	22.21	31.85	34.44	84.88
<i>kappa</i>								79.27

Table H.5. Examinee Classification: Control Condition vs. Study Condition 4 Percentages and Kappa, Sample Size = 1,000

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.06	0.00	0.00	0.00	0.00	0.06
	0+	0.00	0.12	0.10	0.00	0.00	0.00	0.22
	1	0.00	0.00	4.06	0.00	0.00	0.00	4.06
	1+	0.00	0.00	0.36	5.29	0.00	0.00	5.64
	2	0.00	0.00	0.00	1.53	14.37	0.00	15.89
	2+	0.00	0.00	0.00	0.00	7.85	17.54	25.38
	3	0.00	0.00	0.00	0.00	0.00	14.31	34.44
Total	0.00	0.18	4.51	6.82	22.21	31.85	34.44	75.80
<i>kappa</i>								65.92

Table H.6. Examinee Classification: Control Condition vs. Study Condition 5 Percentages and Kappa, Sample Size = 1,000

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.16	0.01	0.00	0.00	0.00	0.17
	1	0.00	0.02	4.47	0.26	0.00	0.00	4.74
	1+	0.00	0.00	0.04	6.49	0.46	0.00	6.98
	2	0.00	0.00	0.00	0.06	20.80	0.00	20.86
	2+	0.00	0.00	0.00	0.00	0.96	26.04	27.00
	3	0.00	0.00	0.00	0.00	0.00	5.81	34.44
Total	0.00	0.18	4.51	6.82	22.21	31.85	34.44	92.40
<i>kappa</i>								89.47

Table H.7. Examinee Classification: Control Condition vs. Study Condition 0 Percentages and Kappa, Sample Size = 200

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.15	0.02	0.00	0.00	0.00	0.17
	1	0.00	0.05	4.13	0.26	0.00	0.00	4.43
	1+	0.00	0.00	0.33	6.32	0.46	0.00	7.10
	2	0.00	0.00	0.00	0.52	19.34	0.22	20.08
	2+	0.00	0.00	0.00	0.00	2.18	25.14	27.32
	3	0.00	0.00	0.00	0.00	0.00	6.40	34.51
Total	0.00	0.20	4.47	7.10	21.97	31.76	34.51	89.59
<i>kappa</i>								85.52

Table H.8. Examinee Classification: Control Condition vs. Study Condition 1 Percentages and Kappa, Sample Size = 200

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.20	0.57	0.00	0.00	0.00	0.76
	1	0.00	0.00	3.90	2.20	0.00	0.00	6.10
	1+	0.00	0.00	0.00	4.86	3.04	0.00	7.90
	2	0.00	0.00	0.00	0.04	17.80	0.62	18.46
	2+	0.00	0.00	0.00	0.00	1.13	22.97	24.10
	3	0.00	0.00	0.00	0.00	0.00	8.18	34.51
Total	0.00	0.20	4.47	7.10	21.97	31.76	34.51	84.23
<i>kappa</i>								78.28

Table H.9. Examinee Classification: Control Condition vs. Study Condition 2 Percentages and Kappa, Sample Size = 200

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.20	0.80	0.00	0.00	0.00	0.99
	1	0.00	0.00	3.67	2.01	0.00	0.00	5.68
	1+	0.00	0.00	0.00	5.04	1.78	0.00	6.82
	2	0.00	0.00	0.00	0.06	17.18	0.10	17.33
	2+	0.00	0.00	0.00	0.00	3.02	21.96	24.98
	3	0.00	0.00	0.00	0.00	0.00	9.71	34.51
Total	0.00	0.20	4.47	7.10	21.97	31.76	34.51	82.55
<i>kappa</i>								76.80

Table H.10. Examinee Classification: Control Condition vs. Study Condition 3 Percentages and Kappa, Sample Size = 200

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.20	0.66	0.00	0.00	0.00	0.85
	1	0.00	0.00	3.81	2.57	0.00	0.00	6.38
	1+	0.00	0.00	0.00	4.54	3.64	0.00	8.18
	2	0.00	0.00	0.00	0.00	17.20	0.69	17.89
	2+	0.00	0.00	0.00	0.00	1.13	22.89	24.02
	3	0.00	0.00	0.00	0.00	0.00	8.18	34.51
Total	0.00	0.20	4.47	7.10	21.97	31.76	34.51	83.14
<i>kappa</i>								76.82

Table H.11. Examinee Classification: Control Condition vs. Study Condition 4 Percentages and Kappa, Sample Size = 200

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.07	0.00	0.00	0.00	0.00	0.07
	0+	0.00	0.12	0.13	0.00	0.00	0.00	0.25
	1	0.00	0.01	3.94	0.10	0.00	0.00	4.05
	1+	0.00	0.00	0.40	5.40	0.01	0.00	5.81
	2	0.00	0.00	0.00	1.60	14.41	0.00	16.01
	2+	0.00	0.00	0.00	0.00	7.56	16.93	24.49
	3	0.00	0.00	0.00	0.00	0.00	14.84	34.51
Total	0.00	0.20	4.47	7.10	21.97	31.76	34.51	75.30
<i>kappa</i>								65.22

Table H.12. Examinee Classification: Control Condition vs. Study Condition 5 Percentages and Kappa, Sample Size = 200

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.18	0.13	0.00	0.00	0.00	0.30
	1	0.00	0.02	4.23	0.90	0.00	0.00	5.15
	1+	0.00	0.00	0.11	5.91	0.94	0.00	6.96
	2	0.00	0.00	0.00	0.30	17.72	0.07	18.09
	2+	0.00	0.00	0.00	0.00	3.31	22.48	25.79
	3	0.00	0.00	0.00	0.00	0.00	9.22	34.51
Total	0.00	0.20	4.47	7.10	21.97	31.76	34.51	85.02
<i>kappa</i>								79.15

Table H.13. Examinee Classification: Control Condition vs. Study Condition 0 Percentages and Kappa, Sample Size = 150

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.15	0.03	0.00	0.00	0.00	0.18
	1	0.00	0.05	4.09	0.31	0.00	0.00	4.45
	1+	0.00	0.00	0.33	6.03	0.43	0.00	6.79
	2	0.00	0.00	0.00	0.74	18.89	0.28	19.91
	2+	0.00	0.00	0.00	0.00	2.56	25.01	27.57
	3	0.00	0.00	0.00	0.00	0.00	6.70	34.39
Total	0.00	0.21	4.45	7.08	21.88	31.99	34.39	88.58
<i>kappa</i>								84.10

Table H.14. Examinee Classification: Control Condition vs. Study Condition 1 Percentages and Kappa, Sample Size = 150

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.20	0.55	0.00	0.00	0.00	0.75
	1	0.00	0.01	3.87	2.17	0.00	0.00	6.05
	1+	0.00	0.00	0.03	4.85	2.92	0.00	7.80
	2	0.00	0.00	0.00	0.05	17.57	0.51	18.13
	2+	0.00	0.00	0.00	0.00	1.39	22.50	23.89
	3	0.00	0.00	0.00	0.00	0.00	8.98	34.39
Total	0.00	0.21	4.45	7.08	21.88	31.99	34.39	83.38
<i>kappa</i>								77.06

Table H.15. Examinee Classification: Control Condition vs. Study Condition 2 Percentages and Kappa, Sample Size = 150

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.21	0.81	0.00	0.00	0.00	1.02
	1	0.00	0.00	3.63	1.95	0.00	0.00	5.58
	1+	0.00	0.00	0.00	5.07	1.69	0.00	6.76
	2	0.00	0.00	0.00	0.07	16.81	0.07	16.94
	2+	0.00	0.00	0.00	0.00	3.38	21.45	24.83
	3	0.00	0.00	0.00	0.00	0.00	10.48	34.39
Total	0.00	0.21	4.45	7.08	21.88	31.99	34.39	81.55
<i>kappa</i>								74.38

Table H.16. Examinee Classification: Control Condition vs. Study Condition 3 Percentages and Kappa, Sample Size = 150

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.21	0.67	0.00	0.00	0.00	0.88
	1	0.00	0.00	3.75	2.39	0.00	0.00	6.13
	1+	0.00	0.00	0.03	4.69	3.09	0.00	7.81
	2	0.00	0.00	0.00	0.01	17.47	0.72	18.19
	2+	0.00	0.00	0.00	0.00	1.32	22.29	23.61
	3	0.00	0.00	0.00	0.00	0.00	8.98	34.39
Total	0.00	0.21	4.45	7.08	21.88	31.99	34.39	82.79
<i>kappa</i>								76.28

Table H.17. Examinee Classification: Control Condition vs. Study Condition 4 Percentages and Kappa, Sample Size = 150

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.05	0.00	0.00	0.00	0.00	0.05
	0+	0.00	0.15	0.13	0.00	0.00	0.00	0.28
	1	0.00	0.01	3.87	0.10	0.00	0.00	3.97
	1+	0.00	0.00	0.45	5.21	0.01	0.00	5.67
	2	0.00	0.00	0.00	1.77	14.41	0.00	16.18
	2+	0.00	0.00	0.00	0.00	7.46	17.09	24.55
	3	0.00	0.00	0.00	0.00	0.00	14.90	34.39
Total	0.00	0.21	4.45	7.08	21.88	31.99	34.39	75.11
<i>kappa</i>								64.91

Table H.18. Examinee Classification: Control Condition vs. Study Condition 5 Percentages and Kappa, Sample Size = 150

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.20	0.24	0.00	0.00	0.00	0.44
	1	0.00	0.01	4.11	1.16	0.00	0.00	5.27
	1+	0.00	0.00	0.10	5.64	1.09	0.00	6.83
	2	0.00	0.00	0.00	0.28	17.28	0.15	17.71
	2+	0.00	0.00	0.00	0.00	3.51	20.94	24.45
	3	0.00	0.00	0.00	0.00	0.00	10.90	34.39
Total	0.00	0.21	4.45	7.08	21.88	31.99	34.39	82.56
<i>kappa</i>								75.72

Table H.19. Examinee Classification: Control Condition vs. Study Condition 0 Percentages and Kappa, Sample Size = 100

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.15	0.02	0.00	0.00	0.00	0.17
	1	0.00	0.12	4.17	0.21	0.00	0.00	4.50
	1+	0.00	0.00	0.45	5.81	0.43	0.00	6.68
	2	0.00	0.00	0.00	1.10	17.97	0.17	19.23
	2+	0.00	0.00	0.00	0.00	3.68	24.68	28.36
	3	0.00	0.00	0.00	0.00	0.00	7.70	34.41
Total	0.00	0.27	4.64	7.12	22.07	32.55	34.41	87.19
<i>kappa</i>								82.03

Table H.20. Examinee Classification: Control Condition vs. Study Condition 1 Percentages and Kappa, Sample Size = 100

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.26	0.48	0.00	0.00	0.00	0.73
	1	0.00	0.01	4.10	1.97	0.00	0.00	6.07
	1+	0.00	0.00	0.06	5.10	2.57	0.00	7.73
	2	0.00	0.00	0.00	0.05	17.28	0.76	18.09
	2+	0.00	0.00	0.00	0.00	2.22	21.65	23.87
	3	0.00	0.00	0.00	0.00	0.00	10.15	34.41
Total	0.00	0.27	4.64	7.12	22.07	32.55	34.41	82.79
<i>kappa</i>								76.12

Table H.21. Examinee Classification: Control Condition vs. Study Condition 2 Percentages and Kappa, Sample Size = 100

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.27	0.76	0.00	0.00	0.00	1.02
	1	0.00	0.00	3.88	1.88	0.00	0.00	5.77
	1+	0.00	0.00	0.00	4.94	1.54	0.00	6.48
	2	0.00	0.00	0.00	0.30	16.22	0.18	16.70
	2+	0.00	0.00	0.00	0.00	4.31	20.12	24.43
	3	0.00	0.00	0.00	0.00	0.00	12.26	34.41
Total	0.00	0.27	4.64	7.12	22.07	32.55	34.41	79.84
<i>kappa</i>								71.81

Table H.22. Examinee Classification: Control Condition vs. Study Condition 3 Percentages and Kappa, Sample Size = 100

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.26	0.52	0.00	0.00	0.00	0.78
	1	0.00	0.01	4.07	2.15	0.00	0.00	6.23
	1+	0.00	0.00	0.04	4.91	2.76	0.00	7.71
	2	0.00	0.00	0.00	0.05	17.32	0.82	18.19
	2+	0.00	0.00	0.00	0.00	2.00	21.68	23.68
	3	0.00	0.00	0.00	0.00	0.00	10.05	34.41
Total	0.00	0.27	4.64	7.12	22.07	32.55	34.41	82.66
<i>kappa</i>								75.96

Table H.23. Examinee Classification: Control Condition vs. Study Condition 4 Percentages and Kappa, Sample Size = 100

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.06	0.00	0.00	0.00	0.00	0.06
	0+	0.00	0.19	0.16	0.00	0.00	0.00	0.35
	1	0.00	0.01	3.86	0.11	0.00	0.00	3.98
	1+	0.00	0.00	0.62	5.13	0.05	0.00	5.80
	2	0.00	0.00	0.00	1.88	13.83	0.00	15.71
	2+	0.00	0.00	0.00	0.00	8.19	16.66	24.85
	3	0.00	0.00	0.00	0.00	0.00	15.89	34.41
Total	0.00	0.27	4.64	7.12	22.07	32.55	34.41	74.09
<i>kappa</i>								63.30

Table H.24. Examinee Classification: Control Condition vs. Study Condition 5 Percentages and Kappa, Sample Size = 100

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.26	0.34	0.00	0.00	0.00	0.60
	1	0.00	0.01	4.15	1.34	0.00	0.00	5.50
	1+	0.00	0.00	0.15	5.24	1.29	0.00	6.68
	2	0.00	0.00	0.00	0.53	15.78	0.14	16.45
	2+	0.00	0.00	0.00	0.00	5.01	18.45	23.46
	3	0.00	0.00	0.00	0.00	0.00	13.97	34.41
Total	0.00	0.27	4.64	7.12	22.07	32.55	34.41	78.29
<i>kappa</i>								69.57

Table H.25. Examinee Classification: Control Condition vs. Study Condition 0 Percentages and Kappa, Sample Size = 75

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.15	0.03	0.00	0.00	0.00	0.18
	1	0.00	0.15	4.06	0.30	0.00	0.00	4.51
	1+	0.00	0.00	0.70	5.74	0.55	0.00	7.00
	2	0.00	0.00	0.00	1.38	16.93	0.67	18.98
	2+	0.00	0.00	0.00	0.00	4.66	23.00	27.75
	3	0.00	0.00	0.00	0.00	0.00	8.46	34.25
Total	0.00	0.30	4.79	7.42	22.14	32.13	34.34	84.12
<i>kappa</i>								77.74

Table H.26. Examinee Classification: Control Condition vs. Study Condition 1 Percentages and Kappa, Sample Size = 75

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.27	0.40	0.03	0.00	0.00	0.70
	1	0.00	0.03	4.18	1.98	0.04	0.00	6.23
	1+	0.00	0.00	0.21	5.00	2.43	0.00	7.64
	2	0.00	0.00	0.00	0.40	16.24	0.87	17.51
	2+	0.00	0.00	0.00	0.00	3.43	20.37	23.96
	3	0.00	0.00	0.00	0.00	0.00	10.89	34.19
Total	0.00	0.30	4.79	7.42	22.14	32.13	34.34	80.25
<i>kappa</i>								72.54

Table H.27. Examinee Classification: Control Condition vs. Study Condition 2 Percentages and Kappa, Sample Size = 75

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.30	0.64	0.00	0.00	0.00	0.94
	1	0.00	0.00	4.10	1.83	0.00	0.00	5.93
	1+	0.00	0.00	0.04	5.12	1.59	0.00	6.76
	2	0.00	0.00	0.00	0.46	15.21	0.42	16.09
	2+	0.00	0.00	0.00	0.00	5.35	19.06	24.40
	3	0.00	0.00	0.00	0.00	0.00	12.66	34.34
Total	0.00	0.30	4.79	7.42	22.14	32.13	34.34	78.13
<i>kappa</i>								69.38

Table H.28. Examinee Classification: Control Condition vs. Study Condition 3 Percentages and Kappa, Sample Size = 75

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	0.28	0.42	0.03	0.00	0.00	0.00	0.73
1	0.00	0.02	4.25	2.04	0.04	0.00	0.00	6.35
1+	0.00	0.00	0.12	4.96	2.58	0.00	0.00	7.66
2	0.00	0.00	0.00	0.39	16.36	0.87	0.00	17.62
2+	0.00	0.00	0.00	0.00	3.16	20.66	0.15	23.97
3	0.00	0.00	0.00	0.00	0.00	10.61	34.19	44.79
Total	0.00	0.30	4.79	7.42	22.14	32.13	34.34	80.70
<i>kappa</i>								73.17

Table H.29. Examinee Classification: Control Condition vs. Study Condition 4 Percentages and Kappa, Sample Size = 75

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
0	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.06
0+	0.00	0.18	0.19	0.00	0.00	0.00	0.00	0.37
1	0.00	0.06	3.84	0.09	0.00	0.00	0.00	3.99
1+	0.00	0.00	0.76	5.26	0.12	0.00	0.00	6.14
2	0.00	0.00	0.00	2.07	13.51	0.00	0.00	15.58
2+	0.00	0.00	0.00	0.00	8.51	15.91	0.00	24.42
3	0.00	0.00	0.00	0.00	0.00	16.22	34.34	50.56
Total	0.00	0.30	4.79	7.42	22.14	32.13	34.34	73.03
<i>kappa</i>								61.82

Table H.30. Examinee Classification: Control Condition vs. Study Condition 5 Percentages and Kappa, Sample Size = 75

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	0.28	0.42	0.00	0.00	0.00	0.00	0.70
1	0.00	0.02	4.16	1.66	0.00	0.00	0.00	5.84
1+	0.00	0.00	0.21	5.06	1.53	0.00	0.00	6.80
2	0.00	0.00	0.00	0.69	14.74	0.30	0.00	15.73
2+	0.00	0.00	0.00	0.00	5.87	16.96	0.00	22.83
3	0.00	0.00	0.00	0.00	0.00	14.88	34.34	49.21
Total	0.00	0.30	4.79	7.42	22.14	32.13	34.34	75.55
<i>kappa</i>								65.74

Table H.31. Examinee Classification: Control Condition vs. Study Condition 0 Percentages and Kappa, Sample Size = 50

Study Condition ILR Level	Control Condition ILR Level							Total
	0	0+	1	1+	2	2+	3	
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.18
1	0.00	0.21	4.81	0.00	0.00	0.00	0.00	5.02
1+	0.00	0.00	1.12	5.09	0.21	0.00	0.00	6.42
2	0.00	0.00	0.00	2.74	15.47	0.00	0.00	18.21
2+	0.00	0.00	0.00	0.00	6.88	18.84	0.00	25.72
3	0.00	0.00	0.00	0.00	0.00	11.44	33.02	44.46
Total	0.00	0.39	5.93	7.82	22.56	30.28	33.02	77.40
<i>kappa</i>								68.61

Table H.32. Examinee Classification: Control Condition vs. Study Condition 1 Percentages and Kappa, Sample Size = 50

Study Condition ILR Level	Control Condition ILR Level							Total
	0	0+	1	1+	2	2+	3	
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	0.25	0.21	0.00	0.00	0.00	0.00	0.46
1	0.00	0.14	5.37	0.98	0.00	0.00	0.00	6.49
1+	0.00	0.00	0.35	5.82	1.37	0.00	0.00	7.54
2	0.00	0.00	0.00	1.02	15.68	0.07	0.00	16.77
2+	0.00	0.00	0.00	0.00	5.51	15.58	0.00	21.09
3	0.00	0.00	0.00	0.00	0.00	14.63	33.02	47.65
Total	0.00	0.39	5.93	7.82	22.56	30.28	33.02	75.72
<i>kappa</i>								66.51

Table H.33. Examinee Classification: Control Condition vs. Study Condition 2 Percentages and Kappa, Sample Size = 50

Study Condition ILR Level	Control Condition ILR Level							Total
	0	0+	1	1+	2	2+	3	
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	0.39	0.49	0.00	0.00	0.00	0.00	0.88
1	0.00	0.00	5.16	1.09	0.00	0.00	0.00	6.25
1+	0.00	0.00	0.28	5.72	0.98	0.00	0.00	6.98
2	0.00	0.00	0.00	1.02	14.42	0.07	0.00	15.51
2+	0.00	0.00	0.00	0.00	7.16	14.84	0.00	22.00
3	0.00	0.00	0.00	0.00	0.00	15.37	33.02	48.39
Total	0.00	0.39	5.93	7.82	22.56	30.28	33.02	73.54
<i>kappa</i>								63.36

Table H.34. Examinee Classification: Control Condition vs. Study Condition 3 Percentages and Kappa, Sample Size = 50

Study Condition ILR Level	Control Condition ILR Level							Total
	0	0+	1	1+	2	2+	3	
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	0.25	0.25	0.00	0.00	0.00	0.00	0.49
1	0.00	0.14	5.37	1.23	0.00	0.00	0.00	6.74
1+	0.00	0.00	0.32	5.79	1.44	0.00	0.00	7.54
2	0.00	0.00	0.00	0.81	15.82	0.07	0.00	16.70
2+	0.00	0.00	0.00	0.00	5.30	15.58	0.00	20.88
3	0.00	0.00	0.00	0.00	0.00	14.63	33.02	47.65
Total	0.00	0.39	5.93	7.82	22.56	30.28	33.02	75.82
<i>kappa</i>								66.69

Table H.35. Examinee Classification: Control Condition vs. Study Condition 4 Percentages and Kappa, Sample Size = 50

Study Condition ILR Level	Control Condition ILR Level							Total
	0	0+	1	1+	2	2+	3	
0	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.04
0+	0.00	0.18	0.04	0.00	0.00	0.00	0.00	0.21
1	0.00	0.18	4.67	0.00	0.00	0.00	0.00	4.84
1+	0.00	0.00	1.23	4.35	0.04	0.00	0.00	5.61
2	0.00	0.00	0.00	3.47	12.46	0.00	0.00	15.93
2+	0.00	0.00	0.00	0.00	10.07	12.39	0.00	22.46
3	0.00	0.00	0.00	0.00	0.00	17.89	33.02	50.91
Total	0.00	0.39	5.93	7.82	22.56	30.28	33.02	67.05
<i>kappa</i>								54.05

Table H.36. Examinee Classification: Control Condition vs. Study Condition 5 Percentages and Kappa, Sample Size = 50

Study Condition ILR Level	Control Condition ILR Level							Total
	0	0+	1	1+	2	2+	3	
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	0.35	0.49	0.00	0.00	0.00	0.00	0.84
1	0.00	0.04	5.23	1.23	0.00	0.00	0.00	6.49
1+	0.00	0.00	0.21	5.65	1.05	0.00	0.00	6.91
2	0.00	0.00	0.00	0.95	13.33	0.00	0.00	14.28
2+	0.00	0.00	0.00	0.00	8.18	12.07	0.00	20.25
3	0.00	0.00	0.00	0.00	0.00	18.21	33.02	51.23
Total	0.00	0.39	5.93	7.82	22.56	30.28	33.02	69.65
<i>kappa</i>								58.04

Table H.37. Examinee Classification: Control Condition vs. Study Condition 0 Percentages and Kappa, Sample Size = 25

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.33	0.00	0.00	0.00	0.00	0.33
	1	0.00	1.67	4.33	0.33	0.00	0.00	6.33
	1+	0.00	0.00	3.00	3.00	0.00	0.00	6.00
	2	0.00	0.00	0.00	4.00	12.67	0.00	16.67
	2+	0.00	0.00	0.00	0.00	8.67	17.33	26.00
	3	0.00	0.00	0.00	0.00	0.00	12.33	32.33
Total	0.00	2.00	7.33	7.33	21.33	29.67	32.33	70.00
<i>kappa</i>								57.81

Table H.38. Examinee Classification: Control Condition vs. Study Condition 1 Percentages and Kappa, Sample Size = 25

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	0.67	0.67	0.00	0.00	0.00	1.33
	1	0.00	1.33	5.67	0.67	0.00	0.00	7.67
	1+	0.00	0.00	1.00	4.00	1.00	0.00	6.00
	2	0.00	0.00	0.00	2.67	12.33	3.00	18.00
	2+	0.00	0.00	0.00	0.00	8.00	10.67	19.33
	3	0.00	0.00	0.00	0.00	0.00	16.00	31.67
Total	0.00	2.00	7.33	7.33	21.33	29.67	32.33	65.00
<i>kappa</i>								51.42

Table H.39. Examinee Classification: Control Condition vs. Study Condition 2 Percentages and Kappa, Sample Size = 25

Study Condition ILR Level	Control Condition ILR Level							
	0	0+	1	1+	2	2+	3	Total
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0+	0.00	2.00	1.00	0.00	0.00	0.00	3.00
	1	0.00	0.00	6.33	0.67	0.00	0.00	7.00
	1+	0.00	0.00	0.00	4.67	0.33	0.00	5.00
	2	0.00	0.00	0.00	2.00	11.33	1.00	14.33
	2+	0.00	0.00	0.00	0.00	9.67	9.33	19.00
	3	0.00	0.00	0.00	0.00	0.00	19.33	32.33
Total	0.00	2.00	7.33	7.33	21.33	29.67	32.33	66.00
<i>kappa</i>								52.46

Table H.40. Examinee Classification: Control Condition vs. Study Condition 3 Percentages and Kappa, Sample Size = 25

Study Condition ILR Level	Control Condition ILR Level							Total
	0	0+	1	1+	2	2+	3	
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	0.67	0.67	0.00	0.00	0.00	0.00	1.33
1	0.00	1.33	6.33	0.67	0.00	0.00	0.00	8.33
1+	0.00	0.00	0.33	4.33	1.00	0.00	0.00	5.67
2	0.00	0.00	0.00	2.33	12.33	3.00	0.00	17.67
2+	0.00	0.00	0.00	0.00	8.00	10.67	0.67	19.33
3	0.00	0.00	0.00	0.00	0.00	16.00	31.67	47.67
Total	0.00	2.00	7.33	7.33	21.33	29.67	32.33	66.00
<i>kappa</i>								52.80

Table H.41. Examinee Classification: Control Condition vs. Study Condition 4 Percentages and Kappa, Sample Size = 25

Study Condition ILR Level	Control Condition ILR Level							Total
	0	0+	1	1+	2	2+	3	
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.67
1	0.00	1.33	5.00	0.00	0.00	0.00	0.00	6.33
1+	0.00	0.00	2.33	3.33	0.00	0.00	0.00	5.67
2	0.00	0.00	0.00	4.00	11.00	0.00	0.00	15.00
2+	0.00	0.00	0.00	0.00	10.33	11.33	0.00	21.67
3	0.00	0.00	0.00	0.00	0.00	18.33	32.33	50.67
Total	0.00	2.00	7.33	7.33	21.33	29.67	32.33	63.67
<i>kappa</i>								49.60

Table H.42. Examinee Classification: Control Condition vs. Study Condition 5 Percentages and Kappa, Sample Size = 25

Study Condition ILR Level	Control Condition ILR Level							Total
	0	0+	1	1+	2	2+	3	
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0+	0.00	1.33	0.67	0.00	0.00	0.00	0.00	2.00
1	0.00	0.67	6.67	1.00	0.00	0.00	0.00	8.33
1+	0.00	0.00	0.00	4.67	0.00	0.00	0.00	4.67
2	0.00	0.00	0.00	1.67	9.67	0.00	0.00	11.33
2+	0.00	0.00	0.00	0.00	11.67	5.33	0.00	17.00
3	0.00	0.00	0.00	0.00	0.00	24.33	32.33	56.67
Total	0.00	2.00	7.33	7.33	21.33	29.67	32.33	60.00
<i>kappa</i>								45.09