DOCUMENT RESUME

ED 478 079 TM 035 020

AUTHOR	Wang, Shudong; Wang, Ning
TITLE	The Effects of Multidimensional Polytomous Response Data on Unidimensional Many-FACET Rasch Model Parameter Estimates.
PUB DATE	2003-04-00
NOTE	24p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Estimation (Mathematics); *Item Response Theory; Simulation
IDENTIFIERS	FACETS Computer Program; Multidimensionality (Tests); *Polytomous Items; Rasch Model; *Unidimensionality (Tests)

ABSTRACT

When categorical responses were simulated from a Multidimensional Many-FACETS Rasch Compensatory Model (MMFRCM), the effects of ability, task difficulty, and step difficulty estimates with the unidimensional Many-FACETS Rasch Model (MFRM; Linacre, 1999) were examined in terms of three error indexes, average absolute difference (AAD), bias, and root mean square error (RMSE). The results show that violating unidimensional assumptions does have an effect on parameter estimation. However, the degree to which estimation shows robustness or not varies dramatically. The conclusion is that the complex nature of the model and data must be clearly understood to determine under which conditions the model should be applied and how well the parameters associated with the model can be estimated reliably. This study provides strong evidence that indicates the nature of MFRM performance when model assumption is violated.(Contains 11 tables and 44 references.) (Author/SLD)

The Effects of Multidimensional Polytomous Response Data on Unidimensional Many-FACET Rasch Model Parameter Estimates

Shudong Wang
Harcourt Educational Measurement

Ning Wang
Promissor

Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 2003. All correspondence should be sent to Shudong Wang, Harcourt Educational Measurement, 19500 Bulverde Road, San Antonio, Texas 78259-3701. E-mail: Shudong_wang@Harcourt.com

Abstract

When categorical responses are simulated from a Multidimensional Many-FACETS Rasch Compensatory Model (MMFRCM), the effects of ability, task difficulty, and step difficulty estimates with unidimensional Many-FACETS Rasch Model (MFRM, Linacre, 1989) were examined in terms of three error indexes, average absolute difference (AAD), bias, and root mean square error (RMSE). The results show that violating unidimensional assumptions do have an effect on parameters estimation. However, the degree to which parameters under which condition that estimation shows robustness or not varies dramatically. The conclusion is that complex nature of the model and data must be clearly understood to determine under which conditions the model should be applied and how well the parameters associated with model can be reliably estimated. This study provides strong evidences which indicates the nature of MFRM performance when model assumption is violated.

The Effects of Multidimensional Polytomous Response Data on Unidimensional Many-FACET Rasch Model Parameter Estimates

Perspective

Essay questions and performance tasks are becoming more important and commonplace in large-scale assessments, such as *Stanford Achievement Test* from Harcourt Educational Measurement), *Scholastic Aptitude Test* (SAT) from ETS and the *TerraNova* from CTB-McGraw Hill. However, essay questions and performance tasks are not without their drawbacks because of the expense, time requirements, and issues of subjectivity associated with scoring. Both human rater and automated-scoring methods in large-scale, high-stakes standardized assessments could cause concerns over validity and fairness of scoring because raters' judgments are treated as the only criteria of essay or performance quality (Bennett & Bejar, 1998; Linacre, 1989; Keith, 1998; Mzumara, Shermis, & Fogel, 1998; Powers, Burstein, Chodorow, Fowles, & Kukich, 2000, 2001).

One of solutions to prescribe ordinal rating observations being ordered qualitatively on latent trait of interest is to use Many-FACETS Rasch Model (MFRM, Linacre, 1989). MFRM is an extension of the partial credit model (Master, 1982) and is a powerful tool to construct linear, objective measures with known precision and quality. MFRM extends the possibility of objective measurement to examinations which include subjective judgments. MFRM also yields greater freedom from judge bias and greater generalizability of the resulting examinee measures than has previously been available (Linacre, 1989). MFRM has been used to conduct analysis on rater behavior, pattern of rating in varied performance assessment situations, and job analysis (Engelhard, 1992, 1994, 1996; Engelhard, Myford, & Cline, 2000; Linacre, Englhard, Tatum, & Myford, 1994; Lumley & McNamara, 1995; Lunz & Stahl, 1990; Myford & Cline, 2002; Wang, 2002). One of the fundamental assumptions about MFRM and many other IRT models is that the variable to be measured is unidimensional. In practice, this assumption of unidimensionality has been violated in most testing situations, and testing professionals now agree that tests are seldom unidimensional (Ackerman, 1992, 1996; Hambleton & Swaminathan, 1985; Reckase, 1979, 1985, 1997; Stout, 1987; Traub, 1983; Yen, 1984, 1985). Using a unidimensional IRT model for multidimensional test data might cause lack of fit of the data to the model; jeopardize "sample-free", "test-free", and "judge-free" properties of the model; and lead to incorrect conclusions about the nature of the data being investigated (Ackerman, 1994; Li & Robert, 2000; Linacre, 1989; Reckase, 1985). Although there were extensive studies on applying unidimensional IRT models to multidimensional tests on other IRT models (Ackerman, 1989; Ansley & Forsyth, 1985; De Ayala, 1994; Drasgow & Parsons, 1983; Folk & Green, 1989; Harrison, 1986; Luecht & Miller, 1992; Oshima & Miller, 1990; Reckase, 1979, 1987; Way, Ansley, & Forsyth, 1986; Kirisci, Hsu, & Yu, 2000), no attempt has been made to directly assess the robustness of violating assumption of unidimensionality on MFRM with polytomous response data. Given the fact that the MFRM was widely used in many situations to address important issues in many fields, the consequence of violation unidimensionality using MFRM should not be continually neglected.

The purpose of this empirical study is to examine the consequences of ability, task difficulty, and step difficulty estimates with the unidimensional MFRM when categorical responses are simulated from a Multidimensional Many-FACETS Rasch Compensatory Model (MMFRCM) and to attempt to provide some understanding of the nature of the ability, task difficulty, and step difficulty estimations under violation of unidimensionality.

The Multidimensional Many-FACETS Rasch Compensatory Model

First, the Multidimensional Many-FACETS Rasch Compensatory Model (MMFRCM) was developed. The MMFRCM is a multidimensional extension of the MFRM (Linacre, 1989). As the distinction was made between compensatory and noncompensatory for the three-parameter logistic model (Ansley & Forsyth, 1985; Hattie, 1981; Simpson, 1978), for all examinees dimensions, the MMFRCM specifies a single task difficulty parameter for each task, a single rater (called "scale" in job analysis) severity/leniency for each rater, and the same set of step difficulties for rating categories (rating category holds across task but differs among rater/scale). The exponential form of the MMFRCM is

$$P(\theta_{nijkh}) = \frac{\exp \sum_{h=1}^{r} \sum_{x=0}^{k} \left[\theta_{nh} - \delta_i - \lambda_j - \tau_{jx}\right]}{\sum_{c}^{K} \exp \sum_{h=1}^{r} \sum_{x=0}^{c} \left[\theta_{nh} - \delta_i - \lambda_j - \tau_{jx}\right]}, k = 0,1,\ldots,K \qquad (1)$$

Where

P is the probability of examinee n for dimension h on task i being rated by rater j, a rating of category k,

θ_{nh} is the ability parameter for examinee n for dimension h (n from 1 to N; h from 1 to r),

δ_i is the difficulty parameter for task i (i from 1 to I),

λ_j is the severity parameter for rater j (j from 1 to J),

τ_{jx} is the step difficulty parameter on rating scale of k categories and for this study, rating category holds across task but differs among rater/scale (x from 0 to K).

Method and Data

Design

To examine the effects of multidimensional polytomous response data on the MFRM parameter estimates, five factors were manipulated and two or three levels of each of the factors were selected. There were 4 independent variables: (1) Ability dimension (one, two, and three), (2) Sample size of examinee (500, 1000, 2000), (3) Degree of ability correlation (0, .3, and .7), (4) Task (40 and 80), (5) Rater/scale (one, two, and three). For two raters/scales, the same 5 step difficulties are -.2, -.05, .05, .2. For three raters/scales, first two raters/scales have same 5 step difficulties: -2, -.05, .05, .2 and third rater/scale has step difficulties -1.5, 0, 1.5. Three error indexes, average absolute difference (AAD), bias, and root mean square error (RMSE) were used as dependent variables for evaluating the effect of the simulation. For the purpose of comparison, responses from a unidimensional MFRM were also generated in the study. Five replications of each of the (1 one dimension + 3 two dimension + 4 three dimension) x 3 sample size x 3 degree of correlation x 2 number of task = 144 total combination (cells) were run. Based on a past research suggestion (Harwell, Stone, Hsu, & Kirisci, 1996), both descriptive and inferential procedures were used to summarize the simulation results.

Simulation procedure.

Given parameters defined by the specifications mentioned above, the steps involved in this simulation process are:

Step 1, a sample of 500, 1000, and 2000 vectors of true abilities were generated from a multivariate normal distribution with specified intercorrectons (2D: $\rho_{12} = 0$, .3, and .7; 3D: $\rho_{123} =$ (0,0,0), (0,0, .3), (0,0, .7), and (0, .3, .7)) using Cholesky factorization procedure (Timm, 1997). For unidimension, same size of samples true ability were generated from standard normal distribution.

Step 2, the known parameters (θ, δ, λ, and τ) were used to calculate the probability of each simulated examinee for each dimension on each task rated by each rater with each a rating of category k using equation (1).

Step 3, the generated probabilities from step 2 were compared to a uniform (0,1) random number to produce responses to specific categories.

The different random numbers were used as seed for each of five replications.

Results

The parameter estimates based on the responses from step 3 were calibrated using FACETS computer program (Linacre, 1996, 1998). For ability, the unidimensional estimates of ability were correlated with both the individual and average true ability parameters, SE and RMSE were calculated. For task, the unidimensional estimates of task difficulties were correlated with both the individual and average true ability parameters, SE and RMSE were calculated.

Ability Estimation

Tables 1 to 5 show the means and standard deviations (SD) of AAD, bias, and RMSE of ability estimations for unidimension, two dimension, and three dimension conditions. These results suggest that, in general, as dimension increases and number of tasks decrease, the AAD, bias, and RMSE of ability estimations increase. The AAD, bias, and RMSE of ability estimation between individual true ability and estimate are larger than those of ability estimations between average true ability and estimates.

Average (over replication and number rater/scale) correlations between estimated ability $\hat{\theta}$ and first true θ_1, second true θ_2, and average true θ_{avg} abilities for two dimensional data are presented in Table 6. For the unidimensional data set, the correlation between true and estimated ability is higher than that of two dimensional data. As correlation $\rho(\theta_1, \theta_2)$ between true abilities increased, the correlation of $r_{\hat{\theta}\theta_1}$ increased too, but this is not necessarily true for $r_{\hat{\theta}\theta_2}$. The $\hat{\theta}$ values were highly related to the averages of the true θs only when the values of $\rho(\theta_1, \theta_2)$ were 0 and 0.3.

Table 7 shows the results of the three-way ANOVA of AAD, bias, and RMSE (averaged across replication) for unidimensional data set. The three factors of number rater (NR or scale), sample size (SS), and number task (NT) have different effects on AAD, bias, and RMSE of ability estimations. None of two two-factor interactions nor the one three-factor interaction effect are statistically significant. The main effects of NR on AAD and bias are statistically significant at 0.05 level. The effect of NT is statistically significant. The NR has the most influence on AAD and bias - it accounted for 9% of the total variance of AAD, and 25.3% of the total variance of bias.

Tables 8 and 9 present the three-way ANOVA of AAD, bias, and RMSE (averaged across replication) for two- and three-dimensional data sets. Three factors manipulated were correlation between true abilities, number rater (or scale), and sample size. For two dimensional data, all interaction effects are not statistically significant. Although the main effect of factor of correlation is statistically significant for RMSE, this factor practically has no effect on ability estimation because it has low values of η^2 that explained percentage variance on total variance. For three dimensional data, some interaction effects are statistically significant but had very low values of η^2. The main effects of factor of correlation are statistically significant for AAD, bias, and RMSE, but it accounted for very low values of the total variance.

Task and Step Estimations

Tables 11 and 12 show the correlations between task estimates and true task parameters under different conditions. First, the number task has no effect on the means and SDs of average correlations. The only effect is number rater (or scale). However, this decrease is due to the number of steps used in factor of number raters. When numbers of one and two raters are used, the number steps is five, added one more rater used 3 steps instead of 5. Although the confounding between number rater and number step could be explained as the contribution to the changes in the values of correlation of task estimation, the real factor should be the number steps rather than the number rater because there is no correlation difference between one rater and two raters.

Practical Implication

This empirical study is the first study to systematically examine the effects of the unidimensional parameter estimates derived from two- and three-dimensional data when the Many-FACETS Rasch Model is used. It seems that violating unidimensional assumptions does have an effect on parameter estimation. However, the degree to which parameters under which condition that estimation shows robustness or not varies dramatically. For this study, among all factors, the number of raters had the most effect on AAD, Bias, and RMSE, and the sample size has least effect on AAD, bias, and RMSE. The number of step and the number of task have moderate effects on AAD, bias, and RMSE. Given the fact that the MFRM is widely used in education, psychological, health, and licensure and certification assessments, the complex nature of the model and data must be clearly understood to determine under which conditions the model should be applied and how well the parameters associated with model can be reliably estimated. This study provides strong evidence which indicates the nature of MFRM performance when model assumption is violated.

Reference

Ackerman, T. (1989). unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement, 13,* 113-127.

Ackerman, T. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement, 29*(1), 67-91.

Ackerman, T. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Psychological Measurement, 18,* 255-278.

Ackerman, T. (1996). Graphical representation of multidimensional item response theory. *Applied Psychological Measurement, 20*, 311-329.

Ansley, T. M., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement, 9,* 39-48.

Bennett, R. E., & Bejar, I. I. (1998). Validity and automated scoring: It's not only the scoring. *Educational Measurement: Issues and Practices, 17*(4), 9-17.

Engelhard, G., Jr. (1992). The measurement of writing ability with a many-faceted Rasch model. *Applied Measurement in Education, 5*, 171-191.

De Ayala, R. J. (1992). The influence of dimensionality on CAT ability estimation. *Educational and Psychological Measurement, 52,* 513-528.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement, 7,* 189-199.

Engelhard, G., Jr. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement, 31,* 93-112.

Engelhard, G. Jr. (1996). Examining rater accuracy in performance assessments. *Journal of Educational Measurement, 33(1),* 56-70.

Engelhard, G., Jr., Myford, C. M., & Cline, F. (2000). *Investigating assessor effects in national Board for Professional Teaching Standards assessments for Early Childhood/Generalist and Middle Childhood/Generalist certification* (ETS Research Report RR-00-13). Princeton, NJ: Educational Testing Service.

Folk, V. G., & Green, B. F. (1989). Adaptive estimation when the unidimensionality assumption of IRT is violated. *Applied Psychological Measurement, 13,* 371-389.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications.* Boston: Kluwer.

Harwell, M. R., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement, 20,* 101-125.

Hattie, J. (1981). *Decision criteria for determining unidimensionality.* Unpublished doctoral dissertation, University of Toronto, Canada.

Kirisci, L., Hsu, T.C., & Yu, L. (2000). Robustness of Selected Calibration Programs to Violations of Assumptions of Unidimensionality of Test Items and Normality of Ability Distribution. *Applied Psychological Measurement, 25*(2), 146-162.

Li, Y. H., & Robert, W. L. (2000). An evaluation of the accuracy of multidimensional IRT linking. *Applied Psychological Measurement, 24*(2), 115-138.

Linacre, J. M. (1989). *Many-faceted Rasch measurement.* Chicago, IL: MESA Press.

Linacre, J. M. (1996). Facets, Version 2.99 [Computer program]. Chicago, IL: MESA Press, University of Chicago.

Linacre, J. M. (1998). *A user's guide to Facets Rasch measurement computer program [Computer program manual].* Chicago, IL: MESA Press, University of Chicago.

Linacre, J. M., Engelhard, G., Jr., Tatum, D. S., & Myford, C. M. (1994). Measurement with judges: Many-faceted conjoint measurement. *International Journal of Educational*

Research, 21(6), 569-577.

Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing, 12*(1), 54-71.

Lunz, M. E., & Stahl, J. A. (1990). Judge consistency and severity across grading periods. *Evaluation and the Health Professions, 13*(14), 425-444.

Masters, G.N. (1982). A Rasch Model for partial credit scoring. *Psychometrika, 47*, 149-174.

Myford, C. M., & Cline, F. (2002). Looking for patterns in disagreements: A facets analysis of human raters' and e-rater's™ scores on essays written for the Graduate Management Adminssion Test® (GMAT®).

Mzumara, H. R., Shermis, M. D., & Fogel, M. (1998). *Validity of the IUPUI placement test scores for course placement: 1997-1998.* Indianapolis: Indiana University Purdue University
Indianapolis (IUPUI) Testing Center.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2000). *Comparing the validity of automated and human essay scoring* (GRE Board Research Report No. 98-08aR). Princeton, NJ: Educational Testing Service.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2001). *Stumping E-Rater: Challenging the validity of automated essay scoring* (GRE Board Research Report No. 98-08bR). Princeton, NJ: Educational Testing Service.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4,* 207-230.

Reckase, M. D. (1985). The difficulty of items that measure more than one ability. *Applied psychological Measurement, 9,* 401-412.

Reckase, M. D. (1987). *A comparison of the results of applying several different unidimensional IRT estimation procedures to multidimensional item response data.* Paper presented at the annual meeting of AERA, Washington, DC.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied psychological Measurement, 21*(1), 25-36.

Simpson, J. B. (1971). A model for testing with multidimensional items. In D. J. Weiss (Ed), *Proceedings of the 1977 Computerized Adaptive Testing Conference* (pp. 82-98). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Stout, W. F. (1987). A nonparametric approach to assessing latent trait dimensionality. *Psychometrika, 52,* 589-617.

Timm, N. H., & Mieczkowski, T. A. (1997). *Univeriate & multivariate general linear models: Theory and applications using SAS® software.* Cary, NC: SAS Institute Inc.

Traub, R. E. (1983). A prior considerations in choosing an item response theory model. In R. K. Hambleton (Ed), *Applications of item response theory, pp. 57-70.* British Columbia, Canada:
Educational Research Institute of British Columbia.

Wang, N. (2002). *Obtaining task importance weights from job analysis survey data: An application of the multi-faceted model.* Paper presented at the 11th biennial meeting of the International Objective Measurement, New Orleans, LA.

Way, W. D., Ansley, T. N., & Forsyth, R. A. (1986). *The effects of two-dimensional data on unidimensional IRT parameter estimates.* Paper presented at the annual meeting of AERA, San Francisco, CA.

Wright, B. D. (1997). *History of Social Science Measurement.* Chicago, IL: MESA Press, University of Chicago.

Yen, W. M. (1984). Effect of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied psychological Measurement, 8,* 125-145.

Yen, W. M. (1985). Increasing item complexity: A possible cause of scale shrinkage for unidimensional item response theory. *Psychometrika, 50,* 399-410.

Table 1

Means and Standard Deviations (over replication) of the AAD, Bias, and RMSE of Ability Parameter Estimations Based on Unidimensional Data

Dimension	No. Step	Sample Size	No. Task	AAD		Bias		RMSE	
				Mean	SD	Mean	SD	Mean	SD
1	5	500	40	.49	.39	.25	.21	.58	.84
			80	.41	.40	.12	.20	.49	.81
		1000	40	.49	.38	.24	.21	.57	.83
			80	.41	.40	.11	.20	.49	.81
		2000	40	.49	.41	.24	.21	.56	.81
			80	.41	.45	.11	.20	.48	.79

Table 2

Means and Standard Deviations (over replication) of the AAD, Bias, and RMSE of Ability Parameter Estimations Based on Two Dimensional Data (All estimates were compared to first true ability)

Correlation $\rho(\theta_1, \theta_2)$	No. Step	Sample Size	No. Task	AAD		Bias		RMSE	
				Mean	SD	Mean	SD	Mean	SD
0	5	500	40	1.00	.41	.45	.35	1.63	.98
			80	.92	.45	.22	.34	1.40	.86
		1000	40	1.08	.41	.63	.60	1.85	1.15
			80	.92	.45	.21	.34	1.39	.83
		2000	40	1.00	.92	.43	.36	1.61	.94
			80	.91	.96	.20	.34	1.38	.83
.3	5	500	40	.99	.91	.41	.36	1.61	1.14
			80	.92	.96	.18	.34	1.47	1.10
		1000	40	.99	.92	.41	.36	1.64	1.13
			80	.91	.97	.18	.33	1.43	1.09
		2000	40	1.00	.92	.42	.35	1.65	1.13
			80	.91	.95	.19	.33	1.43	1.08
.7	5	500	40	.97	.91	.43	.36	1.59	1.33
			80	.88	.95	.21	.32	1.38	1.29
		1000	40	1.04	.91	.63	.59	1.83	1.50
			80	.88	.95	.21	.32	1.43	1.36
		2000	40	1.05	.88	.65	.62	1.84	1.52
			80	.89	.93	.22	.32	1.41	1.32

Table 3

Means and Standard Deviations (over replication) of the AAD, Bias, and RMSE of Ability Parameter Estimations Based on Two Dimensional Data (All estimates were compared to average true ability)

Correlation $\rho(\theta_1, \theta_2)$	No. Step	Sample Size	No. Task	AAD		Bias		RMSE	
				Mean	SD	Mean	SD	Mean	SD
0	5	500	40	.82	.39	.44	.37	1.13	.98
			80	.71	.33	.20	.34	.89	.85
		1000	40	.91	.46	.65	.60	1.35	1.16
			80	.71	.34	.22	.34	.88	.84
		2000	40	.82	.38	.44	.36	1.12	.95
			80	.71	.34	.21	.34	.88	.84
.3	5	500	40	.87	.40	.42	.36	1.28	1.13
			80	.79	.37	.18	.34	1.12	1.07
		1000	40	.87	.39	.44	.36	1.29	1.12
			80	.76	.36	.21	.33	1.07	1.06
		2000	40	.87	.40	.43	.35	1.29	1.13
			80	.77	.37	.20	.33	1.07	1.07
.7	5	500	40	.92	.42	.42	.36	1.46	1.33
			80	.82	.41	.20	.32	1.25	1.29
		1000	40	.99	.48	.63	.59	1.69	1.51
			80	.83	.41	.21	.32	1.28	1.37
		2000	40	1.00	.49	.64	.62	1.70	1.52
			80	.83	.41	.21	.32	1.27	1.33

Table 4

Means and Standard Deviations (over replication) of the AAD, Bias, and RMSE of Ability Parameter Estimations Based on Three Dimensional Data (All estimates were compared to first true ability)

Correlation $\rho(\theta_1, \theta_2, \theta_3)$	No. Step	Sample Size	No. Task	AAD		Bias		RMSE	
				Mean	SD	Mean	SD	Mean	SD
(0, 0, 0)	3 and 5	500	40	1.20	1.21	.57	.46	2.27	1.23
			80	1.08	1.25	.26	.42	1.89	.92
		1000	40	1.23	1.17	.57	.48	2.38	1.27
			80	1.10	1.22	.25	.43	1.97	.92
		2000	40	1.21	1.08	.55	.48	2.31	1.23
			80	1.10	1.14	.24	.43	1.93	.90
(0, 0, .3)	3 and 5	500	40	1.22	1.20	.54	.49	2.25	.85
			80	1.09	1.26	.23	.43	1.84	.40
		1000	40	1.29	1.18	.57	.47	2.61	1.10
			80	1.19	1.24	.26	.42	2.26	.88
		2000	40	1.28	1.09	.52	.49	2.58	1.11
			80	1.21	1.15	.28	.40	2.34	.93
(0, 0, .7)	3 and 5	500	40	1.28	1.12	.58	.45	2.61	1.49
			80	1.16	1.17	.29	.41	2.24	1.40
		1000	40	1.29	1.12	.52	.46	2.65	1.46
			80	1.18	1.16	.22	.42	2.31	1.40
		2000	40	1.33	1.16	.53	.46	2.85	1.40
			80	1.20	1.22	.24	.42	2.47	1.41
(0, .3, .7)	3 and 5	500	40	1.20	.90	.53	.45	2.33	1.53
			80	1.11	.96	.24	.41	2.07	1.49
		1000	40	1.22	.91	.50	.46	2.44	1.61
			80	1.12	.96	.24	.40	2.16	1.57
		2000	40	1.22	1.11	.55	.42	2.41	1.56
			80	1.12	1.15	.25	.41	2.16	1.54

Table 5

Means and Standard Deviations (over replication) of the AAD, Bias, and RMSE of Ability Parameter Estimations Based on Three Dimensional Data (All estimates were compared to average true ability)

Correlation $\rho(\theta_1, \theta_2, \theta_3)$	No. Step	Sample Size	No. Task	AAD		Bias		RMSE	
				Mean	SD	Mean	SD	Mean	SD
(0, 0, 0)	3 and 5	500	40	1.02	.44	.55	.48	1.65	1.23
			80	.87	.34	.25	.43	1.26	.89
		1000	40	1.04	.46	.59	.49	1.71	1.30
			80	.88	.35	.27	.43	1.28	.92
		2000	40	1.03	.45	.57	.48	1.67	1.26
			80	.88	.35	.26	.43	1.28	.90
(0, 0, .3)	3 and 5	500	40	1.02	.41	.55	.48	1.62	.95
			80	.86	.28	.25	.42	1.19	.59
		1000	40	1.12	.43	.57	.47	2.01	1.24
			80	1.00	.37	.26	.42	1.66	1.04
		2000	40	1.05	.43	.51	.50	1.73	1.22
			80	.96	.36	.27	.40	1.52	1.03
(0, 0, .7)	3 and 5	500	40	1.23	.37	.57	.45	2.37	1.38
			80	1.13	.33	.28	.41	2.04	1.26
		1000	40	1.25	.35	.53	.46	2.44	1.31
			80	1.14	.31	.23	.42	2.10	1.23
		2000	40	1.28	.35	.53	.46	2.56	1.33
			80	1.13	.43	.24	.42	2.20	1.35
(0, .3, .7)	3 and 5	500	40	1.15	.43	.52	.45	2.12	1.45
			80	1.04	.41	.23	.41	1.85	1.40
		1000	40	1.17	.43	.52	.46	2.21	1.53
			80	1.05	.41	.26	.40	1.92	1.48
		2000	40	1.16	.43	.55	.43	2.18	1.49
			80	1.05	.42	.26	.41	1.91	1.48

Table 6. Average Correlations among $\hat{\theta}$ and θ_1, θ_2, and Average θ_{avg} for One- and Two-Dimensional Conditions

Dimension	$\rho(\theta_1, \theta_2)$	Sample Size	$r_{\hat{\theta}, \theta_1}$		$r_{\hat{\theta}, \theta_2}$		$r_{\hat{\theta}, \theta_{avg}}$	
			Mean	SD	Mean	SD	Mean	SD
1	-	500	.99	.01				
		1000	.99	.01				
		2000	.99	.01				
2	0	500	.58	.17	.75	.04	.99	.01
		1000	.69	.01	.75	.06	.99	.01
		2000	.70	.01	.65	.04	.99	.01
	.3	500	.80	.01	.80	0	.99	0
		1000	.80	0	.81	0	.99	0
		2000	.80	.01	.71	0	.99	0
	.7	500	.91	0	.31	0	.76	0
		1000	.91	0	.34	.01	.77	0
		2000	.91	0	.33	.01	.77	0

Table 7

Results of ANOVA for Unidimensional Data

		AAD			Bias			RMSE		
Source	DF	F	p	η^2	F	p	η^2	F	p	η^2
Main Effects										
No. Rater (NR)	2	3.961	0.023	0.098	13.995	0.000	0.253	0.615	0.543	0.017
Sample Size (SS)	2	0.000	1.000	0.000	0.033	0.968	0.001	0.002	0.998	0.000
No. Task (NT)	1	0.907	0.344	0.011	10.176	0.002	0.092	0.190	0.664	0.003
Interaction Effects										
NS x SS	4	0.000	1.000	0.000	0.004	1.000	0.000	0.000	1.000	0.000
NR x NT	2	0.001	0.999	0.000	0.146	0.864	0.003	0.016	0.984	0.000
SS x NT	2	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000
SS x NR x NT	4	0.000	1.000	0.000	0.001	1.000	0.000	0.000	1.000	0.000
Error	72									

Table 8

Results of ANOVA for Two Dimensional Data Based on Estimates and First True Ability

		AAD			Bias			RMSE		
Source	DF	F	p	η^2	F	p	η^2	F	p	η^2
Main Effects										
Correlation (C)	2	0.084	0.919	0.001	1.225	0.296	0.008	463.465	0.000	0.000
No. Rater (NR)	2	0.160	0.852	0.001	16.673	0.000	0.115	0.040	0.961	0.001
Sample Size (SS)	2	0.111	0.895	0.001	0.499	0.608	0.003	0.132	0.876	0.001
Interaction Effects										
C x NR	4	0.331	0.857	0.005	0.737	0.567	0.010	0.104	0.901	0.006
C x SS	4	0.074	0.990	0.001	0.343	0.849	0.005	0.356	0.840	0.001
NS x SS	4	0.065	0.992	0.001	0.649	0.628	0.009	0.071	0.991	0.001
C x NR x SS	8	0.071	1.000	0.002	0.308	0.962	0.009	0.057	0.994	0.002
Error	243									

Table 9

Results of ANOVA for Three Dimensional Data Based on Estimates and First True Ability

		AAD			Bias			RMSE		
Source	DF	*F*	*p*	η^2	*F*	*p*	η^2	*F*	*p*	η^2
Main Effects										
Correlation (C)	2	4846.558	0.000	0.012	372.466	0.000	0.000	1210.414	0.000	0.013
No. Scale (NR)	2	1.477	0.221	0.002	0.037	0.990	0.189	1.535	0.205	0.001
Sample Size (SS)	2	0.344	0.709	0.027	45.387	0.000	0.105	0.164	0.849	0.018
Interaction Effects										
C x NR	4	9.897	0.002	0.008	50.478	0.000	0.000	6.458	0.011	0.010
C x SS	4	0.472	0.829	0.000	0.011	1.000	0.000	0.604	0.727	0.000
NR x SS	4	0.028	0.994	0.029	0.031	0.993	0.005	0.041	0.989	0.018
C x NR x SS	8	5.302	0.005	0.000	1.114	0.329	0.000	3.150	0.044	0.000
Error	243									

Table 10

Means and Standard Deviations (over replication) of Average Correlations between Task Estimate and True Task Difficulty for Two Dimensional Data

Dimension	Correlation $\rho(\theta_1, \theta_2)$	No. Rater	No. Task	r Mean	r SD
1	-	1	40	1.00	.00
			80	1.00	.00
		2	40	1.00	.00
			80	1.00	.00
		3	40	.97	.01
			80	.97	.01
2	0	1	40	1.00	.00
			80	1.00	.00
		2	40	1.00	.00
			80	1.00	.00
		3	40	.96	.01
			80	.96	.01
	.3	1	40	1.00	.00
			80	1.00	.00
		2	40	1.00	.00
			80	1.00	.00
		3	40	.97	.01
			80	.96	.01
	.7	1	40	1.00	.00
			80	1.00	.00
		2	40	1.00	.00
			80	1.00	.00
		3	40	.97	.01
			80	.97	.01

Table 11

Means and Standard Deviations (over replication) of Average Correlations between Task Estimate and True Task Difficulty for Three Dimensional Data

Correlation $\rho(\theta_1, \theta_2, \theta_3)$	No. Rater	No. Task	r Mean	r SD
(0, 0, 0)	1	40	1.00	.00
		80	1.00	.00
	2	40	1.00	.00
		80	1.00	.00
	3	40	.96	.01
		80	.96	.01
(0, 0, .3)	1	40	1.00	.00
		80	1.00	.00
	2	40	1.00	.00
		80	1.00	.00
	3	40	.96	.01
		80	.96	.01
(.7, 0, 0)	1	40	1.00	.00
		80	1.00	.00
	2	40	1.00	.00
		80	1.00	.00
	3	40	.96	.01
		80	.96	.01
(.3, .7, .3)	1	40	1.00	.00
		80	1.00	.00
	2	40	1.00	.00
		80	1.00	.00
	3	40	.96	.01
		80	.96	.01

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: The Effects of Multidimensional Polytomous Response Data on Unidimensional Many-FACET Rasch Model Parameter Estimates	
Author(s): Shudong Wang, Ning Wang	
Corporate Source: Harcourt Educational Measurement Promissor	Publication Date: 4/2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☒ Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Level 2A ☐ Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Level 2B ☐ Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Shudong Wang	Printed Name/Position/Title: Shudong Wang, Ph.D.	
Organization/Address: Harcourt Educational Measurement 19500 Bulverde Road San Antonio, Texas 78259-3701	Telephone: 210-339-5535	FAX:
	E-Mail Address: Shudong_wang@Harcourt.com	Date: 9/29/03

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: info@ericfac.piccard.csc.com
WWW: http://ericfacility.org