Open Journal of Applied Sciences, 2014, 4, 425-432
Published Online July 2014 in SciRes. http://www.scirp.org/journal/ojapps
http://dx.doi.org/10.4236/ojapps.2014.48039

Scientific Research

Effects of Differential Item Discriminations between Individual-Level and Cluster-Level under the Multilevel Item Response Theory Model

Chalie Patarapichayatham, Akihito Kamata

Southern Methodist University, University Park, USA
Email: cpatarapichy@smu.edu

Received 3 May 2014; revised 17 June 2014; accepted 29 June 2014

Abstract

This study attempted to interpret differential item discriminations between individual and cluster levels by focusing on patterns and magnitudes of item discriminations under 2PL multilevel IRT model through a set of variety simulation conditions. The consistency between the mean of individual-level ability estimates and cluster-level ability estimates was evaluated by the correlations between them. As a result, it was found that they were highly correlated if the patterns of item discriminations were the same for both individual and cluster levels. The magnitudes of item discriminations themselves did not affect much on correlations, as far as the patterns were the same at the two levels. However, it was found that the correlation became lower when the patterns of item discriminations were different between the individual and cluster levels. Also, it was revealed that the mean of the estimated individual-level abilities would not be necessarily a good representation of the cluster-level ability, if the patterns were different at the two levels.

Keywords

Multilevel Item Response Theory Model, Ability Estimates, Item Discrimination

1. Introduction

Multilevel modeling has become a popular data analysis technique in psychological and educational measurement. Traditional psychometric models, such as classical test theory and item response theory (IRT) models, do not account for a nested structure of the data. Multilevel modeling becomes important when researchers analyze

How to cite this paper: Patarapichayatham, C. and Kamata, A. (2014) Effects of Differential Item Discriminations between Individual-Level and Cluster-Level under the Multilevel Item Response Theory Model. *Open Journal of Applied Sciences*, **4**, 425-432. http://dx.doi.org/10.4236/ojapps.2014.48039

nested data, because it takes into account of both within and between cluster variations of the data. One of popular multilevel modeling techniques is a hierarchical generalized linear model (HGLM). However, when HGLM is applied to multilevel IRT [1], one limitation is that all item discriminations are assumed to be equal. In other words, the relationships between the observed measurement indicators and the latent factor are assumed to be equal for all items in a test, which is sometimes an unrealistic assumption. If item discriminations are allowed to freely vary, the model may more closely resemble the observed data.

IRT models define the relationship between observed item scores and latent constructs for dichotomous and polytomous item response data. An IRT framework has been extended to a multilevel data structure [1]-[3]. A multilevel IRT model is desirable when item response data have been collected from a sample with a nested data structure. In addition to the benefit of modeling data variations both at between and within cluster levels, relationships between variables at different levels can be estimated better as well.

One popular form of an IRT model for dichotomously scored items is a 2-parameter logistic (2PL) model, where the probability of an individual correctly responding to an item depends on individual's ability, item difficulty, and item discrimination. The 2PL IRT model can be written as

$$P_j\left(\theta_i\right)=\frac{\exp\left[a_j\left(\theta_i-b_j\right)\right]}{1+\exp\left[a_j\left(\theta_i-b_j\right)\right]}, \tag{1}$$

where $P_j\left(\theta_i\right)$ is the probability that individual i with ability θ_i answers item j correctly, θ_i is an ability level for individual i, a_j is an item discrimination parameter for item j, and b_j is an item difficulty parameter for item j. The numerator and the second term in the denominator $\exp\left[a_j\left(\theta_i-b_j\right)\right]$ correspond to the odds ratio of a correct response $y_j\left(\theta_i\right)=1$.

The 2PL IRT model also allows the item discriminations to vary freely across items in a test. Its extension to a multilevel IRT model has been investigated and documented by several authors [4]-[9]. However, not much attention has been given to what it means when patterns and/or magnitudes of item discriminations are different between individual and cluster levels. For example, Fox [4] illustrated the importance of taking measurement errors from different sources into account by a multilevel analysis. Although the author presented a new approach to multilevel modeling with three real data sets of mathematics by using 2PL IRT model, difference in item discriminations between levels were beyond his focus. Fox [5] focused on measuring latent dependent and independent variables of a multilevel model, where manifest variables consisting of binary, ordinal, or graded responses. This extension made it possible to model relationships between observed and latent variables on different levels using dichotomous and polytomous IRT models. However, different item discrimination patterns between levels were not a focus of this study. Natesan [7] studied the accuracy and precision of the item parameter estimates of the 2PL multilevel IRT model by varying test lengths, sample sizes, correlation between the predictor variable and the ability parameter, and the distribution shape of the predictor variables interact to impact the accuracy and the precision. Again, differential item discrimination patters between levels were not a focus of this study.

Some authors investigated a cluster-level IRT modeling. For example, Mislevy [6] proposed a notion of the cluster-level IRT model by exploring the relationships between individual-level and cluster-level IRT models, as well as the parameter estimates under the cluster-level IRT model. Results showed that when item response data are gathered in a design of one item per scale from each individual, it is possible to define a cluster-level IRT model. The cluster-level ability estimate was analogous to the individual-level ability estimates. The cluster-level IRT model parameters specified the probability of a correct response to a given item from an individual selected at random from a given cluster. The cluster-level IRT model parameter estimate was a straightforward generalization of an individual-level IRT model technique. Tate [8] studied whether the cluster-level IRT model was robust to typical violations of distributional assumptions under the two-parameter cluster-level IRT dichotomous model through a simulation study. Results showed that the estimated precision was always either approximately consistent with the actual precision or a conservative estimate of the actual precision. When the items were replaced to target high-ability schools, the conservatism increased as school ability moved away from the target ability range. Also, Tate [9] extended to a similar study with a polytomous model. Results were similar to findings from [8]. It is notable that Tate [9] concluded that the estimate of cluster-level ability for a specific cluster could be viewed as the mean of the individual-level ability of all individuals in that cluster if the individual abilities within each cluster are normally distributed with a mean equal to the cluster-level ability, and

the cluster-level ability is also normally distributed. However, no discussion was provided regarding the effect of differential item discriminations between individual and cluster levels. Our concern was whether the patterns and the magnitudes of the item discrimination between levels would affect the estimates of individual or cluster level abilities differently.

Assuming we fit a 1PL multilevel IRT model. The pattern of item discrimination would be exactly the same for both individual and cluster levels. In other word, it obtains only one pattern of the item discriminations for both levels (e.g., same pattern of item discrimination across both levels). The magnitudes of item discrimination would be exactly the same for both individual and cluster levels (e.g., 1.0 for all items across both levels). On the other hand, once the 2PL IRT model is applied to the multilevel model, the patterns and the magnitudes of the item discrimination could be different across levels. While the item discriminations may have the same patterns and the same magnitudes for both individual and cluster levels, the item discriminations may have the same patterns but different magnitudes between individual and cluster levels. The item discriminations may also have different patterns and different magnitudes between individual and cluster levels. However, effects of the difference in different patterns and magnitudes of item discrimination have not been demonstrated in literature yet with the multilevel IRT modeling perspective.

If differential patterns or magnitudes of item discriminations between levels affect estimates of individual and cluster level abilities differently, the conclusion that Tate [9] has made may not be always correct. Namely, estimated cluster-level ability should not always be viewed as the mean of the individual-level abilities in the cluster. Therefore, the mean individual level abilities may practically over- or under-estimates the cluster-level ability. In practice, it is not uncommon to estimate the cluster-level ability from the mean of the individual-level ability of all individuals in that cluster. For example, it is very common to evaluate school level performance by computing the mean of estimated student abilities in each school. For these reason, we attempted to investigate how the patterns and the magnitudes of the item discrimination between individual and cluster levels would behave under the 2PL multilevel IRT model. It was also our intention to provide some insights on how we should interpret the differential item discriminations between individual and cluster levels.

An item discrimination indicates a quality of an item, because it dictates how strongly each item correlates with the ability being measured. A higher discrimination corresponds to a greater correlation with the ability, which also leads to a higher scoring weight for the item. Therefore, individuals who answer items with high discriminations correctly would have a higher estimated ability than individuals who answer items with lower discriminations correctly, given the same raw scores. In other words, it matters not only how many items were answered correctly, but also which items were answered correctly. Under the 2PL single-level IRT model, the individual ability estimates are weighted by item discriminations, such that

$$\sum_{j}^{N} a_j P\left(\theta_i, a_j, b_j\right) = \sum_{j}^{N} a_j u_j \,, \tag{2}$$

where θ_i is an ability level for individual i, a_j is an item discrimination parameter for item j, b_j is an item difficulty parameter for item j, and u_j takes the values 0 and 1 corresponding to incorrect and correct responses of individual i on item j (j = 1, 2, …, N). However, it may or may not be a case under the 2PL multilevel IRT model.

Since 2PL multilevel IRT model allows item discriminations to vary both at the individual and cluster levels, we hypothesized that different patterns of item discriminations between the levels affect patterns of scoring weights to be different between the levels. For this reason, our hypothesis was that the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ should be very close to the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$, if the patterns of item discriminations are the same between the individual and cluster levels. However, if the patterns of item discriminations are different between the levels, we hypothesized that the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ will be different from the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. We evaluated these hypotheses by computing the Pearson's product moment correlation between the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. Note that literature on cluster-level IRT model [6] [8] [9] has suggested that the use of the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ in lieu of the estimated clus-

ter-level ability $\left(\hat{\theta}_k^{(3)}\right)$ is discouraged, due to inappropriate estimation of its standard errors. However, this study focused on the relationship between the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ with respect to patterns of item discriminations at the individual and cluster levels, as an attempt to provide an insight on how one should interpret differential item discrimination parameters between the levels.

2. Design of the Simulation Study

2.1. Modeling

The 2PL multilevel IRT model was investigated under this study for both data generation and fitting the model. In IRT, the response outcome data are treated as categorical with binomial distributions. The idea of a 2PL multilevel IRT model is to measure each latent variable incorporated in the multilevel model with the IRT model. The 2PL multilevel IRT model can be written as

$$y_{ijk}^{*} = a_i^{(2)}\theta_i^{(2)} + a_k^{(3)}\theta_k^{(3)} - b_j + e_{ijk}, \tag{3}$$

where y_{ijk}^{*} is the latent continuous response variable with the observed response $y_{ijk}=1$ if $y_{ijk}^{*}\geq 0$, or $y_{ijk}=0$ if $y_{ijk}^{*}<0$. On the other hand, $a_i^{(2)}$ is the item discrimination parameter for individual level *i*, $\theta_i^{(2)}$ is the unobserved latent ability level for individual *i*, $a_k^{(3)}$ is the item discrimination parameter for the cluster level *k*, $\theta_k^{(3)}$ is the unobserved latent ability level for cluster *k*, b_j is the item difficulty parameter for item *j*, and e_{ijk} is the error term with the logistic distribution. Both $\theta_i^{(2)}$ and $\theta_k^{(3)}$ were generated from standard normal distribution with the mean of zero and the standard deviation of 1.0.

2.2. Simulation Conditions

It was assumed that the test consisted of 12 items with item difficulties ranging from −2.0 to 2.0. These item difficulties were fixed across conditions. The biserial correlations between the ability estimate and the propensity for correct response were used to set up the item discriminations for individual and cluster levels. These five item discriminations (0.8, 1.0, 1.2, 1.6, and 2.2) represented biserial correlations of 0.40, 0.50, 0.55, 0.66, and 0.77 between the latent trait and the propensity to a correct answer. They created 16 sets of item discriminations classified into three patterns (see **Table 1**). The first pattern (Pattern A) was for conditions with the same patterns and the same magnitudes of discriminations for both levels. The second pattern (Pattern B) was for conditions with the same patterns but different magnitudes of discriminations between the levels, and the third pattern (Pattern C) was for conditions with different patterns and different magnitudes of discriminations between levels. Also, the number of clusters (2 levels; 50 and 100) and cluster size (2 levels; 50 and 100) were manipulated, and these three simulation factors were crossed and created a total of 16 × 2 × 2 = 64 simulation conditions.

Based on these specifications, the dichotomous item response data were randomly generated and fit by the 2PL multilevel IRT model as shown in Equation (3). The M*plus* software, using the Maximum Likelihood estimator with robust standard errors was used to estimate model parameters ($a_i^{(2)}$, $a_k^{(3)}$, and b_j), while assuming e_{ijk} are randomly drawn from the standard logistic distribution. On the other hand, $\theta_i^{(2)}$ and $\theta_k^{(3)}$ were estimated by empirical Bayes estimator, also by the M*plus* software. The variances of the latent variables were fixed to be 1.0 for both individual and cluster levels. This way, the magnitudes of item discriminations were directly comparable. 100 replications were generated for each simulation condition. Then, the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ were estimated. Finally, we further evaluated the consistency of the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ by computing the Pearson's product moment correlation between $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and $\left(\hat{\theta}_k^{(3)}\right)$.

3. Results and Discussions

Results demonstrated three primary scenarios regarding Pearson's product moment correlation between the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ (see **Figure 1**).

Table 1. Item discriminations.

Set	Pattern	Item Discriminations			
		Individual Level		Cluster Level	
		Items 1-6	Items 7-12	Items 1-6	Items 7-12
1	A	0.8		0.8	
2	B	0.8		1.2	
3	B	0.8		2.2	
4	B	1.2		0.8	
5	A	1.2		1.2	
6	B	1.2		2.2	
7	B	2.2		0.8	
8	B	2.2		1.2	
9	A	2.2		2.2	
10	A	1.0	1.6	1.0	1.6
11	C	0.8		1.0	1.6
12	C	1.2		1.0	1.6
13	C	2.2		1.0	1.6
14	C	1.0	1.6	0.8	
15	C	1.0	1.6	1.2	
16	C	1.0	1.6	2.2	

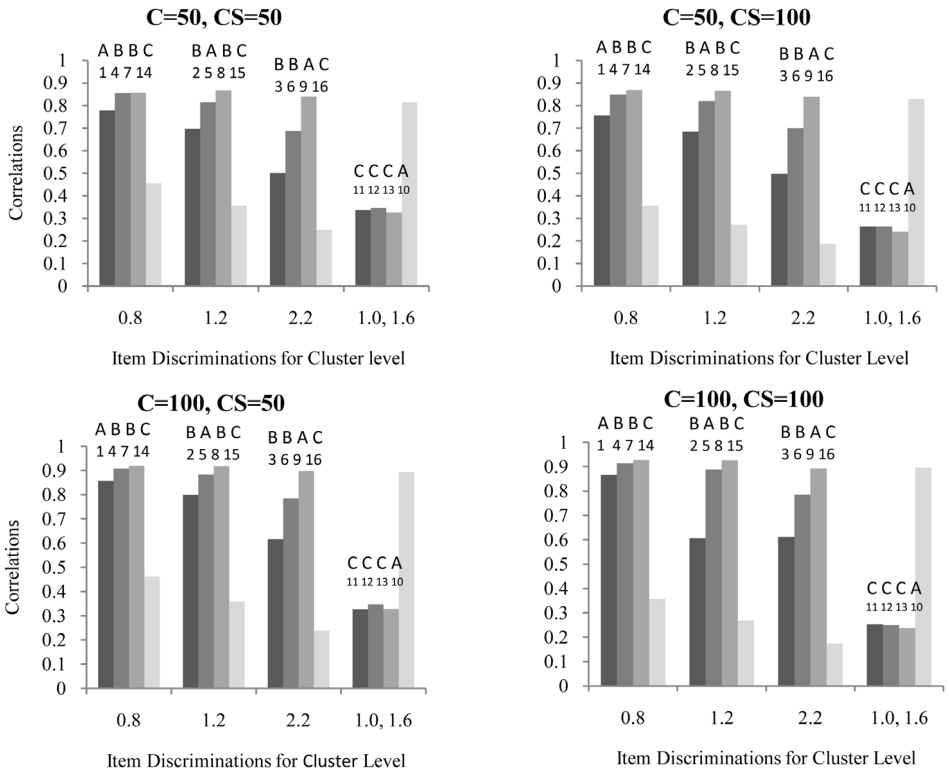
First, conditions with the same patterns and the same magnitudes of item discriminations for both levels (Pattern A), and conditions with the same patterns of item discriminations but the magnitude of item discriminations were higher for individual level (Sets 4, 7, and 8 under Pattern B), the correlations were very high. They were in the range of [0.757, 0.928]. These results indicated that the patterns and the magnitudes of item discriminations between levels affected the estimates and the correlations between the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. In other words, the cluster-level ability could be viewed as the mean of the individual-level ability of all individuals in that cluster. This statement also holds for the 1PL multilevel IRT as a special case. This result suggested that it is reasonable if one estimates the cluster-level ability by computing the mean of the individual-level ability of all individuals in that cluster under these circumstances. Both approaches would produce similar results. Also, within this pattern, the correlations slightly increased as the number of clusters and the cluster size increased.

Second, for conditions with the same patterns of item discriminations between levels but the magnitudes of item discriminations were lower for individual level (Sets 2, 3, and 6 under Pattern B), the correlations were moderately high. They were in the range of [0.497, 0.799]. These results showed that the individual level needed to be at least at the same level of item discriminations for cluster level for the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ to be similar, given the same patterns for both levels. In other words, the item discriminations for individual level played significant role over the item discriminations for cluster level to obtain similar estimates of the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. This result makes sense to us because it is suspected that we lose some information when we aggregate the mean individual-level abilities $\left(\hat{\theta}_i^{(2)}\right)$. Thus, the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ estimates slightly differ from the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$.

Third, for conditions with different patterns and different magnitudes of item discriminations between levels (Pattern C), the correlations were much lower. They were in the range of [0.174, 0.462] among the six Pattern C conditions. In this scenario, it became obvious that the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ were inconsistent. These results confirmed that differential patterns of

item discriminations had a huge impact on the estimates of the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$, whereas the magnitudes of item discriminations did not affect much on the estimates of the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. As can be seen from sets 11, 12, and 13 that the item discriminations for individual level increased from 0.8 to 1.2, and 2.2, however, the correlations for all three sets were almost identical. Similar to the correlations obtained from sets 14. 15. and 16. In other words, the individual level abilities are not the same as the cluster-level abilities, which means that the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ is not a straightforward generalization of the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$. These results confirmed differential item discriminations under the 2PL multilevel IRT model behave differently from the differential item discriminations under the 2PL single-level IRT model.

Overall, the Pearson's product moment correlations between the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ demonstrated that differential patterns of item discriminations between levels had huge impact on the estimates of the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ over the magnitudes of item discriminations between levels. The patterns of item discriminations have to be at least the same between levels to obtain similar estimates between



Note: C = number of cluster, CS = cluster size, A = conditions with the same patterns and the same magnitudes of item discriminations for both levels, B = conditions with the same patterns but different magnitudes of item discriminations between levels, C = conditions with different patterns and different magnitudes of item discriminations between levels, and 1, …, 16 = simulation sets 1 to 16.

Figure 1. Correlations between the aggregated mean individual-level abilities $\left(\bar{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$.

the aggregated mean individual-level abilities $\left(\overline{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. Thus, the individual traits have different meaning than the cluster-level traits under conditions where the patterns and the magnitudes of the item discriminations differ.

Regarding the effect of simulation factors (see also **Figure 1**), for conditions with the same patterns and the same magnitudes of discriminations for both levels (Pattern A) and conditions with the same patterns but different magnitudes of discriminations between the levels (Pattern B), the correlations slightly increased as the number of cluster and the cluster size increased. On the other hand, for conditions with different patterns and different magnitudes of discriminations between levels (Pattern C), the correlations slightly decreased as the number of cluster and the cluster size increased. These results demonstrated that the estimates of the aggregated mean individual-level abilities $\left(\overline{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ were affected by sample sizes; however, not necessarily in the way we typically understand the effect of sample size.

Regarding the effect of item discrimination magnitudes, first, for conditions with the same patterns and the same magnitudes of item discriminations for both levels (Pattern A), the correlations increased as the magnitudes of either item discriminations for the individual level or item discriminations for the cluster level increased. These results were not surprising to us because the item discrimination indicates a quality of an item. A higher item discrimination corresponds to a greater correlation with the ability, which also leads to a higher scoring weight for the item.

Second, for conditions with the same patterns but different magnitudes of discriminations between the levels (Pattern B), the correlations increased as the magnitudes of item discriminations for the individual level increased, given the same magnitudes of item discriminations for the cluster level. On the other hand, the correlations decreased as the magnitudes of item discriminations for the cluster level increased, given the same magnitudes of item discriminations for the individual level. Also, the correlations were much lower when the magnitudes of item discriminations of the individual level were much lower than the magnitudes of item discriminations of the cluster level. Thus, the difference in magnitudes of item discriminations between levels affected the Estimates of the aggregated mean individual-level abilities $\left(\overline{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. This result confirmed that the patterns of item discriminations have to be at least the same between levels to obtain similar estimates between the aggregated mean individual-level abilities $\left(\hat{\theta}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. The magnitudes of item discrimination is related to the magnitude of the variances of latent factor, in our case, the variances of individual-level ability $\left(\theta_i^{(2)}\right)$ and cluster-level ability $\left(\theta_k^{(3)}\right)$. Since latent factors do not have fixed scales, item discriminations do not have fixed scales. Their scales depend on how the latent factors are scaled. Therefore, it makes sense that our results showed that the differential magnitudes of item discriminations did not affect the results, as far as patterns are the same between the levels.

Third, for conditions with different patterns and different magnitudes of item discriminations between levels (Pattern C), the correlations decreased as the magnitudes of either item discriminations for the individual level or item discriminations for the cluster level increased. These findings were interesting, because they were different from what we typically understand the effect of item discrimination; that is, higher item discrimination leads to a higher scoring weight for the item. If this statement is true under the 2PL multilevel IRT modeling, the aggregated mean individual-level abilities $\left(\hat{\theta}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ should be highly correlated, when the magnitudes of either item discriminations for the individual level or item discriminations for the cluster level increased. It is important to note that this statement is always true for a 2PL single-level IRT model, but it is not always the case once we fit the 2PL IRT model under the multilevel modeling (e.g., 2PL multilevel IRT model), as evident from low correlations between the aggregated mean individual-level abilities $\left(\overline{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ when both levels have different patterns of item discriminations and the magnitudes of either item discriminations for the individual level or item discriminations for the cluster level increased. This result also demonstrated that the patterns of the item discriminations had substantial effect on ability estimates under the 2PL multilevel IRT model.

Finally, it should be noted that there were interaction effects between our simulation factors. Namely, the correlations between the aggregated mean individual-level abilities $\left(\overline{\hat{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ slightly increased as the number of clusters, the cluster size, and the item discrimination magnitudes in-

creased under conditions with the same patterns and the same magnitudes of item discriminations for both levels (Pattern A) and for conditions with the same patterns but different magnitudes of discriminations between the levels (Pattern B). On the other hand, the correlations slightly decreased as the number of cluster, the cluster size, and the item discrimination magnitudes increased for conditions with different patterns and different magnitudes of item discriminations between levels (Pattern C).

4. Conclusions

Our investigation demonstrated one way to interpret differential item discriminations between individual and cluster levels in the 2PL multilevel IRT context. We found that the patterns of item discriminations between levels are more important than the magnitudes of item discriminations to obtain consistent estimates of the aggregated mean individual-level abilities $\left(\hat{\bar{\theta}}_i^{(2)}\right)$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. As Mislevy [6] and Tate [8] [9] among others point out, it is not uncommon in educational research to use the aggregated mean individual-level abilities $\left(\hat{\bar{\theta}}_i^{(2)}\right)$, instead of the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$. However, our finding revealed that this would be justifiable only if the magnitudes of item discriminations for both levels share the same patterns. For this reason, a caution should be made when aggregating individually estimated abilities, because the mean of the estimated individual-level abilities $\left(\hat{\bar{\theta}}_i^{(2)}\right)$ is not necessarily a good representation of the cluster level ability $\left(\hat{\theta}_k^{(3)}\right)$.

Although our investigation looked into three simulation factors over 64 simulation conditions, there are of course other factors and conditions that should be investigated. For example, it would be interesting to see how the number of items on each level would affect the estimates of the aggregated mean individual-level abilities $\hat{\bar{\theta}}_i^{(2)}$ and the estimated cluster-level ability $\left(\hat{\theta}_k^{(3)}\right)$ given the same or different patterns of item discriminations between levels. The item difficulty is another interesting factor. It would be interesting to see the effect of the interaction between the item difficulties and the patterns of item discriminations between levels. The extremely low or high item discrimination magnitudes under different patterns of item discriminations between levels should be of interest for future investigation. Also, it would be interesting to further algebraically demonstrate how the individual ability estimates are weighed by the differential item discriminations under the 2PL multilevel IRT model.

References

[1] Kamata, A. (2001) Item Analysis by the Hierarchical Generalized Linear Model. *Journal of Educational Measurement*, **38**, 79-93. http://dx.doi.org/10.1111/j.1745-3984.2001.tb01117.x

[2] Adams, R.J., Wilson, M. and Wu, M. (1997) Multilevel Item Response Models: An Approach to Errors in Variables Regression. *Journal of Educational and Behavioral Statistics*, **22**, 47-76. http://dx.doi.org/10.3102/10769986022001047

[3] Fox, J.P. and Glas, C.A.W. (2001) Bayesian Estimation of a Multi-Level IRT Model Using Gibbs Sampling. *Psychometrika*, **6**6, 271-288. http://dx.doi.org/10.1007/BF02294839

[4] Fox, J.-P. (2004) Applications of Multilevel IRT Modeling. *School Effectiveness and School Improvement*, **15**, 261-280. http://dx.doi.org/10.1080/09243450512331383212

[5] Fox, J.-P. (2005) Multilevel IRT Using Dichotomous and Polytomous Response Data. *British Journal of Mathematical and Statistical Psychology*, **58**, 145-172. http://dx.doi.org/10.1348/000711005X38951

[6] Mislevy, R.J. (1983) Item Response Models for Grouped Data. *Journal of Educational Statistics*, **8**, 271-288. http://dx.doi.org/10.2307/1164913

[7] Natesan, P. (2007) Estimation of Two-Parameter Multilevel Item Response Models with Predictors: Simulation and Substantiation for an Urban School District. Unpublished Doctoral Dissertation, Texas A&M University, College Station, TX.

[8] Tate, R.L. (1995) Robustness of the School-Level IRT Model. *Journal of Educational Measurement*, **32**, 145-162. http://dx.doi.org/10.1111/j.1745-3984.1995.tb00460.x

[9] Tate, R. (2000) Robustness of the School-Level Polytomous IRT Model. *Educational and Psychological Measurement*, **60**, 20-37. http://dx.doi.org/10.1177/00131640021970349

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.