American Journal of Theoretical and Applied Statistics
2022; 11(4): 109-113
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20221104.11
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Effect of Correlation Between Abilities Under Between-Item Dimensionality

Xiong Rao

Department of Statistics and Data Science, University of Arizona, Tucson, United State

Email address:
raoxiong@email.arizona.edu

To cite this article:
Xiong Rao. Effect of Correlation Between Abilities Under Between-Item Dimensionality. *American Journal of Theoretical and Applied Statistics*. Vol. 11, No. 4, 2022, pp. 109-113. doi: 10.11648/j.ajtas.20221104.11

Received: June 16, 2022; **Accepted**: July 22, 2022; **Published**: July 29, 2022

Abstract: The Item Response Theory (IRT) evaluates the relationship between people's ability and test items, and it includes unidimensional and multidimensional models. One key assumption for the unidimensional IRT model is that only one dimension of ability should be tested. However, since people's abilities are latent, many datasets fitted with the unidimensional IRT model reflect abilities from more than one dimension in fact. To identify the consequence of fitting the unidimensional IRT model on correlated abilities, this research focuses on when the correlated abilities can be treated as a single ability, the possible pattern of misfit, and if it is reduced by higher correlated abilities. In the research, the misfits are evaluated by applying unidimensional 2-parameter logistic (2PL) IRT model while the datasets are simulated with items testing two different correlated. The dimensionalities are examined with abilities correlated to different degrees, and the misfit of using the unidimensional IRT model is tested by comparing the item difficulties and item discriminations from the fitted model and the true parameters. The results show that when the correlation between abilities is higher than 0.95, the unidimensional model can be fit without bias. But for all simulated datasets with correlated abilities below 0.95, the estimated item parameters using the unidimensional model are biased and the biases are not reduced with increasing correlation if multiple factors are identified for abilities.

Keywords: The Item Response, People's Ability, The IRT Model, Item Dimensionality, Probability and Statistics, The Factor Analysis

1. Introduction

1.1. The IRT Model

The Item Response Theory (IRT) models are used to discover the relationship between one's unobservable ability, or latent traits, and the observed responses. In the IRT model, the probability of the score one received given his/her ability is a function of his/her ability, characteristic of items, and the possible value for the score [1, 4, 15].

1.2. Dimensionality

IRT models can be classified as unidimensional and multidimensional. In the unidimensional IRT model, people's ability, or latent trait, that brings to answer the test should be unidimensional. This means the items are testing one single skill or knowledge. However, in the multidimensional IRT model (MIRT), there are multiple abilities have been requested for answering the test. [11, 14] If each item reflects only one ability, that is the between-item dimensionality model. In contrast, if one item tests multiple abilities, then there is a within-item dimensionality. [7]

The multidimensional IRT model is more complex than the unidimensional IRT model, but in reality, abilities are usually correlated, so the MIRT model fits most of the cases better. Even though some tests are created to test just one single ability, the truth might be that there is more than one ability being tested.

1.3. Correlated Abilities

If the unidimensional IRT model is used, the estimate of item characters and people's latent traits will absolutely be influenced by putting items and latent traits from different dimensions on the same scale. [10, 13] However, if the abilities tested are not correlated, then a test that has between-item dimensionality could be separately evaluated using the unidimensional IRT model. And another extreme case would be that the abilities are highly correlated, then probably there is indeed only one latent trait measured. [2, 6, 10] The correlation between latent traits plays a big role in fitting models. [3, 8]

When abilities are highly correlated, they can then be treated as a single ability, but how high should the correlation be? And if the unidimensional IRT model is fitted on the items testing multiple abilities, we know that there is bias in estimation, but what's the possible pattern of the bias? Also, would a higher correlation between abilities reduce this bias? With these three questions, this paper focuses on the items with between-item dimensionality and tries to figure out the influence of the correlation between abilities on item characteristic estimators.

2. Method

To examine the effect of correlation between abilities, different data sets of response scores were simulated by manipulating the correlation of abilities. There are 10 items with 5 of them reflecting one ability and the other 5 reflecting another ability. So, a total of two dimensions are examined. The unidimensional IRT model was then used to estimate the item characteristics and then compared to the original item parameters set to simulate the data. [5, 12]

3. Two-Parameter Logistic Model

There are different kinds of IRT models, and the one used in this research is the Two-Parameter Logistic Model (2-PL Model). It is called the 2-PL model because it uses two characteristics of test items, including item difficulty and item discrimination. [1, 15, 16]

3.1. Item Difficulty and Item Discrimination

In the unidimensional IRT model, item difficulty (d), which is also the location of the item characteristic curve (ICC), measures the difficulty level of an item, and it is measured with the scale of latent trait (theta), which is similar with the standard deviation and usually ranges from -3 to +3. The value of difficulty corresponds to the latent trait level that is required to get the item correct with a 50% chance; the smaller value indicates the item is easier, and the bigger value indicates a harder item. For example, if an item has a difficulty level of 2, then people with 2 latent trait levels above the average would have a 50 percent probability to answer the item correctly. The item discrimination (a), the slope of the curve, indicates how effective an item can be used to distinguish people among different levels of the latent trait. Lower discrimination, indicates a flatter slope, corresponding to the smaller change of probability from a latent trait to another latent trait, while the item with higher discrimination separates people into two parts more effectively. [15, 16]

In the MIRT model, the two parameters are different from the unidimensional model since the latent trait is now a vector with multiple abilities (thetas) from different dimensions. In the MIRT model, the ICC is a surface instead of a curve. In this research, there are two abilities from different dimensions, and thus each item has two slopes corresponding to two dimensions. The item reflecting ability 1 would have higher item discrimination on dimension 1, and a lower slope on dimension 2, which is reasonable as the item would hardly discriminate against people on a latent trait if it does not measure that latent trait. If a1 and a2 are two slope parameters, then the multidimensional discrimination equals the sum of two vectors. The d parameter in the MIRT model is not the difficulty parameter, as it cannot indicate the difficulty along. The negative d divided by a slope parameter gives the difficulty level related to that specific dimension and the multidimensional difficulty. [11]

3.2. Parameter Setting

In this research, the slopes have two vectors, a1 and a2, which reflect the two dimensions. The slopes of items for the dimension they reflected are around 1 and are close to 0 for the other dimension. To make them look like real data, values were pulled from normal distribution instead of fixing at 1 and 0. Item 1 to item 5 have higher slopes in a1, and items 6 to 10 have higher slopes in a2. The d parameters were also drawn from a normal distribution and are relatively close to the center. To control the effect of pre-set parameters, the first five slopes in the first column are the same as the last five in the second column, the slopes close to 0 are the same for two dimensions, and the d parameter for the first five items are also the same as the rest five items' parameter. The abilities were set with mean of -.4 and .5, which are close to 0 but not fixed at 0. The variance of theta 1 is 1.21 and the variance of theta 2 is 1.96. The covariance between thetas were set at 0.77, 0.924, 1.078, 1.232, 1.386, and 1.463 to indicate the correlation of 0.5, 0.6, 0.7, 0.8, 0.9 and 0.95. Five datasets with 5000 observations were simulated under each level of correlation.

4. Result

The two-factor model was first applied to each data to check the dimensionality. All datasets with a correlation below 0.95 have a simple structure of loadings with no cross-loading under the 0.3 cut-off, and the loadings are in the range from 0.38 to 0.7. However, for a correlation 0.95, almost all the items are loaded on only one of the factors, and the factor correlations given by the two-factor model are below 0.3. This result shows that when the correlation between abilities reaches 0.95, although the data were generated under two different abilities, the two-factor model is not suitable, which means there is no difference between the two abilities, and thus they can be treated as one. [9]

Although the correlation between thetas was set at 0.5, 0.6, 0.7, 0.8, and 0.9, the factor correlations analyzed from the two-factor model are slightly different. The correlations for simulated data are relatively higher for levels 0.5, and 0.6, and they are lower for levels 0.8 and 0.9. This could probably be limited by the slope parameters. The differences between slopes for two dimensions would affect the loading somehow. For the slope parameters set in this research, there is obviously a gap between the two dimensions, but the gap is only around 1. Thus, the correlation is unlikely to be as high as 0.9 or as low as 0.5.

The estimated item difficulties and discriminations are shown in the table 1 to table 3 at the end of the paper. For the item difficulty, items 1 to 5 have higher estimators compared to the true multidimensional item difficulty, and items 6 to 10 have lower estimators. While parameters and d parameters were controlled for two sets of items, the result should be driven by the distribution of thetas. As the mean of the second theta is 0.9 latent trait level higher than the first theta, the ability that items 6 to 10 tested is generally higher than the ability tested by items 1 to 5. Thus, the difficulty of item 6 to 10 becomes easier while item 1 to 5 becomes harder. However, back to the original hypotheses of the research, compared to the multidimensional item difficulties (B), there is not any pattern showing that datasets from higher ability correlations have a better estimation of item difficulty.

Table 1. *Item Difficulty Table.*

Cor (θ)	**0.5**						**0.6**						**0.7**					
Factor-corerlation	**0.61**	**0.57**	**0.61**	**0.62**	**0.61**	**ave**	**0.71**	**0.69**	**0.7**	**0.65**	**0.68**	**ave**	**0.74**	**0.74**	**0.75**	**0.74**	**0.72**	**ave**
Item 1	-0.14	-0.18	-0.18	-0.19	-0.14	-0.17	-0.15	-0.13	-0.08	-0.1	-0.14	-0.12	-0.18	-0.1	-0.14	-0.16	-0.12	-0.14
Item 2	0.47	0.46	0.47	0.44	0.46	0.46	0.41	0.45	0.44	0.55	0.42	0.46	0.47	0.35	0.43	0.4	0.35	0.4
Item 3	0.23	0.17	0.21	0.19	0.19	0.2	0.16	0.18	0.14	0.21	0.17	0.17	0.14	0.19	0.21	0.17	0.11	0.17
Item 4	0.76	0.85	0.88	0.81	0.76	0.81	0.82	0.78	0.82	0.94	0.76	0.82	0.72	0.76	0.78	0.77	0.71	0.75
Item 5	-0.11	-0.1	-0.04	-0.07	-0.14	-0.09	-0.09	-0.06	-0.01	-0.11	-0.05	-0.06	-0.06	-0.11	-0.06	-0.07	-0.13	-0.08
Item 6	-0.76	-0.76	-0.72	-0.64	-0.66	-0.71	-0.66	-0.64	-0.67	-0.69	-0.68	-0.67	-0.73	-0.7	-0.66	-0.73	-0.67	-0.7
Item 7	-0.23	-0.27	-0.23	-0.22	-0.28	-0.25	-0.23	-0.24	-0.22	-0.22	-0.19	-0.22	-0.27	-0.24	-0.2	-0.27	-0.29	-0.25
Item 8	-0.5	-0.48	-0.47	-0.45	-0.49	-0.48	-0.47	-0.4	-0.4	-0.48	-0.47	-0.44	-0.46	-0.46	-0.45	-0.44	-0.48	-0.46
Item 9	-0.01	0.08	0.02	0.07	0.09	0.05	0.03	0.09	0.06	0.08	0.07	0.07	0.06	0.05	0.09	0.04	0.03	0.05
Item 10	-0.69	-0.68	-0.6	-0.57	-0.58	-0.62	-0.61	-0.59	-0.6	-0.6	-0.63	-0.61	-0.67	-0.6	-0.68	-0.63	-0.64	-0.65

0.8						**0.9**						**MDiff (B)**
0.77	**0.78**	**0.78**	**0.77**	**0.84**	**ave**	**0.78**	**0.81**	**0.81**	**0.81**	**0.84**	**ave**	
-0.1	-0.16	-0.16	-0.13	-0.11	-0.13	-0.13	-0.14	-0.08	-0.15	-0.13	-0.13	-0.5
0.39	0.41	0.4	0.42	0.38	0.4	0.4	0.39	0.45	0.38	0.37	0.4	0.12
0.15	0.13	0.12	0.14	0.12	0.13	0.14	0.16	0.18	0.13	0.12	0.15	-0.18
0.71	0.67	0.7	0.72	0.71	0.7	0.73	0.71	0.74	0.69	0.7	0.71	0.52
-0.1	-0.11	-0.11	-0.04	-0.12	-0.09	-0.09	-0.1	-0.05	-0.09	-0.12	-0.09	-0.43
-0.67	-0.65	-0.69	-0.64	-0.67	-0.66	-0.66	-0.65	-0.63	-0.67	-0.67	-0.66	-0.5
-0.24	-0.24	-0.28	-0.24	-0.26	-0.25	-0.25	-0.25	-0.2	-0.25	-0.24	-0.24	0.12
-0.48	-0.44	-0.49	-0.46	-0.42	-0.46	-0.46	-0.44	-0.41	-0.45	-0.45	-0.44	-0.18
0.04	0.04	0.02	0.02	0.06	0.04	0.04	0.06	0.11	0.01	0.02	0.05	0.52
-0.59	-0.59	-0.59	-0.57	-0.64	-0.6	-0.61	-0.6	-0.58	-0.6	-0.59	-0.59	-0.43

Table 2. *Item Discrimination Table.*

Cor (θ)	**0.5**						**0.6**						**0.7**					
Factor-corerlation	**0.61**	**0.57**	**0.61**	**0.62**	**0.61**	**ave**	**0.71**	**0.69**	**0.7**	**0.65**	**0.68**	**ave**	**0.74**	**0.74**	**0.75**	**0.74**	**0.72**	**ave**
Item_1	1.14	1.05	1.07	1.02	1.09	1.07	1.19	1.14	1.25	1.04	1.19	1.16	1.08	1.22	1.27	1.3	1.21	1.22
Item_2	0.97	0.87	0.87	1.02	0.89	0.92	0.98	0.97	1.04	0.91	0.98	0.98	1	1.13	1.01	1.1	1.05	1.06
Item_3	1.12	1.15	1.03	1.1	1.05	1.09	1.22	1.22	1.26	1.19	1.25	1.23	1.36	1.29	1.31	1.27	1.36	1.32
Item_4	0.87	0.79	0.88	0.84	0.89	0.85	0.93	0.93	0.92	0.87	0.97	0.92	1.04	0.99	0.91	0.86	1.02	0.96
Item_5	0.88	0.84	0.81	0.85	0.89	0.85	0.93	0.91	0.98	0.91	0.93	0.93	1.03	1.09	0.94	0.98	0.95	1
Item_6	1.41	1.48	1.55	1.65	1.69	1.55	1.77	1.77	1.56	1.55	1.61	1.65	1.64	1.51	1.63	1.66	1.7	1.63
Item_7	1.27	1.25	1.33	1.4	1.29	1.31	1.36	1.34	1.28	1.41	1.31	1.34	1.39	1.23	1.43	1.34	1.35	1.35
Item_8	1.71	1.67	1.69	1.76	1.71	1.71	1.73	1.8	1.86	1.81	1.65	1.77	1.68	1.71	1.75	1.85	1.67	1.73
Item_9	1.14	1.04	1.16	1.14	1.18	1.13	1.26	1.2	1.22	1.04	1.24	1.19	1.12	1.11	1.19	1.27	1.2	1.18
Item_10	1.21	1.21	1.24	1.32	1.36	1.27	1.3	1.26	1.26	1.42	1.21	1.29	1.24	1.22	1.24	1.34	1.28	1.26

0.8						**0.9**						**MDisc (A)**
0.77	**0.78**	**0.78**	**0.77**	**0.84**	**ave**	**0.78**	**0.81**	**0.81**	**0.81**	**0.84**	**ave**	
1.23	1.31	1.27	1.26	1.33	1.28	1.33	1.37	1.34	1.32	1.37	1.35	1.23
1.07	1.09	1.06	1.08	1.16	1.09	1.17	1.17	1.17	1.11	1.1	1.14	1
1.29	1.45	1.33	1.26	1.53	1.37	1.41	1.47	1.48	1.46	1.43	1.45	1.34
0.95	1.01	1.02	0.95	1	0.99	0.99	1.02	1.04	1.05	0.99	1.02	0.87
0.98	1.06	1.01	1.03	0.99	1.01	1.05	1.12	1.03	1.07	1.02	1.06	0.92
1.67	1.73	1.8	1.69	1.76	1.73	1.79	1.81	1.77	1.77	1.84	1.8	1.23
1.34	1.44	1.35	1.32	1.34	1.36	1.38	1.45	1.46	1.33	1.43	1.41	1
1.83	1.76	1.82	1.73	1.91	1.81	1.85	1.8	1.87	1.79	1.87	1.84	1.34
1.13	1.14	1.21	1.2	1.33	1.2	1.2	1.22	1.25	1.25	1.28	1.24	0.87
1.3	1.36	1.35	1.38	1.4	1.36	1.3	1.46	1.32	1.27	1.36	1.34	0.92

Table 3. *True-Parameter Table.*

	a1	a2	d	A	B
Item 1	1.22	0.12	0.61	1.23	-0.50
Item 2	0.99	0.10	-0.12	1.00	0.12
Item 3	1.34	0.12	0.24	1.34	-0.18
Item 4	0.86	0.10	-0.45	0.87	0.52
Item 5	0.91	0.11	0.40	0.92	-0.43
Item 6	0.12	1.22	0.61	1.23	-0.50
Item 7	0.10	0.99	-0.12	1.00	0.12
Item 8	0.12	1.34	0.24	1.34	-0.18
Item 9	0.10	0.86	-0.45	0.87	0.52
Item 10	0.11	0.91	0.40	0.92	-0.43

The estimated item discrimination, on the other hand, shows a pattern of decreasing accuracy as the correlation between ability becomes higher, which proves the opposite side of the hypothesis. The item discrimination estimators are closer to the true parameters for items 1 to 5 and are inflated for items 6 to 10. These might be caused by the higher variance of theta in the second dimension. While there is an increasing pattern of estimated discrimination, the estimated parameters for datasets with ability correlation around 0.6 and 0.7 are more accurate than the datasets from other correlation levels. When the abilities are correlated closer, the item discrimination from different dimensions may tend to point in the same direction. Thus, the sum of the two slope parameters, which shows the maximum ability of discrimination, will become higher and higher, since to transform a square into a rhombus, the diagonal becomes longer even if the sides do not change.

5. Conclusion

Therefore, as long as there are multiple factors identified, the estimated item difficulty in the unidimensional model would be biased, and the bias would not be affected by a closer correlation between factors. The item discrimination estimator will be affected by the correlation, but it is not true that the bias is reduced with increasing correlation. For the setting of true parameters in this research, a correlation around 0.6 to 0.7 would be optimal for the smallest bias of item discrimination. Also, the correlation of about 0.95 or above between two abilities shows no difference from a single ability. However, an optimal correlation for the smallest bias of item discrimination and for treating multiple abilities as one would only be applied to this specific case, as there is a limitation from the pre-set parameters. Therefore, the optimal values are different if the a-parameters changed.

In general, factor analysis is required to check if the ability tested is unidimensional. If there are multidimensional abilities, then they're always a bias of estimating people's latent traits using the inaccurate item characteristics. Even though the bias of item discrimination might be small for an optimal correlation between multiple abilities, without knowing the true distributions of abilities, it is hard to predict or measure the bias.

References

[1] Embretson, S. E., & Reise, S. P. (2000). Item Response Theory for Psychologists (1st ed.). Psychology Press. https://doi.org/10.4324/9781410605269

[2] Fayers, P. M., & Hays, R. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In: Fayers P, Hays RD, editors. Assessing quality of life in clinical trials: methods and practice, pp. 55–73. Oxford University Press.

[3] Gosz, J., K. & Walker, C. M. (2002, April). An empirical comparison of multidimensional item response data using TESTFACT and NOHARM. Paper presented at the annual meeting of the National Council for Measurement in Education, New Orleans, LA.

[4] Hambleton, R. K. & Swaminathan, H. (1985). Item response theory: principles and applications. Boston: Hingham, MA, U.S.A: Kluwer-Nijhoff Pub.

[5] Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory (Vol. 2). Sage.

[6] Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and ltems. Applied psychological measurement, 9 (2), 139-164.

[7] Immekus, J. C., & Imbrie, P. K. (2008). Dimensionality Assessment Using the Full-Information Item Bifactor Analysis for Graded Response Data: An Illustration With the State Metacognitive Inventory. Educational and Psychological Measurement, 68 (4), 695–709. https://doi.org/10.1177/0013164407313366

[8] Knol, D. L., & Berger, M. P. F. (1991). Empirical comparison between factor analysis and multidimensional item response models. Multivariate Behavioral Research, 26, 457-477.

[9] Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. Journal of Applied Psychology, 87 (3), 517–529. https://doi.org/10.1037/0021-9010.87.3.517

[10] Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of educational statistics, 4 (3), 207-230.

[11] Reckase, M. D. (2009). Multidimensional Item Response Theory. New York, NY: Springer.

[12] Rindskopf, D., and Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. Multivar. Behav. Res. 23, 51–67. doi: 10.1207/s15327906mbr2301_3.

[13] Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. Psychometrika, 52 (4), 589-617.

[14] Svetina, D., Valdivia, A., Underhill, S., Dai, S., & Wang, X. (2017). Parameter Recovery in Multidimensional Item Response Theory Models Under Complexity and Nonnormality. Applied psychological measurement, 41 (7), 530–544. https://doi.org/10.1177/0146621617707507

[15] Van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). Handbook of modern item response theory. New York, NY: Springer.

[16] Wu, M., Tam, H. P., & Jen, T. H. (2016). Educational measurement for applied researchers. Theory into practice, 136.