SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

Does factor indeterminacy matter in multi-dimensional item response theory?

- Chong Ho Yu, Ph.D., D. Phil

USERS PROGRAM

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Does factor indeterminacy matter in multi-dimensional item response theory?

Chong Ho Yu, Ph.D., D. Phil., cyu@apu.edu

Azusa Pacific University, USA

ABSTRACT

This paper aims to illustrate proper applications of multi-dimensional item response theory (MIRT), which is available in SAS's PROC IRT. MIRT combines item response theory (IRT) modeling and factor analysis when the instrument carries two or more latent traits. While it may seem convenient to accomplish two tasks simultaneously by employing one procedure, users should be cautious of mis-interpretations. This illustration utilizes the 2012 Programme for International Student Assessment (PISA) data set. Because there are two known sub-domains in the PISA test (reading and math), PROC IRT was programmed to adopt a two-factor solution. Additionally, the loading plot, dual plot, item difficulty/discrimination plot, and test information function plot in JMP were utilized to examine the psychometric properties of the PISA test. When reading and math items were analyzed in SAS's MIRT, 7-13 factors are suggested. At first glance these results are puzzling because ideally all items should be loaded into two factors. However, when the psychometric attributes yielded from a 2–parameter IRT analysis are examined, it is evident that both the reading and math test items are well-written. It is concluded that even if factor indeterminacy is present, it is advisable to evaluate its psychometric soundness based on IRT because content validity can supersede construct validity.

INTRODUCTION

- Item response theory (IRT) assume uni-dimensionality. In other words, a test or a survey should examine only a single latent trait of participants. In reality, many tests or surveys are multi-dimensional; to address this issue multi-dimensional IRT (MIRT) was introduced. Besides accounting for multidimensionality, MIRT also aims to model latent covariance structures between multiple dimensions, and to model these interactions (Hartig & Hoher, 2009).
- A classic example is about the latent traits behind a math test. When a math problem is presented in a formula or an equation, then the required problem-solving ability is said to be the mathematical skill alone. However, if the math item is explained in text, then it might require both math and reading skills to solve the problem.
- MIRT is a fusion of factor analysis and IRT. However, the underlying philosophy of IRT is vastly different from that of factor analysis, which belongs to the realm of Classical Test Theory (CTT). For example, the psychometric attributes of an instrument yielded from factor analysis attach to the entire scale whereas each item developed by IRT has its own characteristic (item difficulty parameter, discrimination parameter, guessing parameter, item information function, etc.). Hence, it is legitimate to generate an adaptive test by selecting items from an item bank.

Acknowledgments: Special thanks to Ms. Anna Lee and Samantha Douglas for their valuable input to this article.

METHODOLOGY

Data Source. This illustration utilizes the 2012 Programme for International Student Assessment (PISA) data set collected by Organization for Economic Cooperation and Development (OECD, 2013). In 2012, 51,000 students from 65 countries participated in PISA. In order to rule out cultural difference as an extraneous variable, in this analysis only USA students are selected (n = 4978). However, not all students answered all test items; rather, different booklets were assigned to different students and the missing values were imputed to obtain the plausible value for each student. To simplify the illustration, a subset of US students and a subset of items are selected so that imputation is not necessary. As a result, only 411 students are retained.

One may argue that both math and science require similar reasoning approaches and thus the two factors are indistinguishable in the data is expected. But according to Gardner (2006), linguistic skill and math skill belong to different dimensions of intelligence. To illustrate factor indeterminacy in this assessment, the authors retain reading and math items only.

Data analysis. Because there are two known sub-domains in the test (reading and math), PROC IRT in SAS was forced to adopt a two-factor solution. The loading plot, dual plot, item difficulty/discrimination plot, and test information function plot in JMP were utilized to examine the psychometric properties of the PISA test. The dual plot is a graphic display that places item parameters and student ability (theta) on the same scale, whereas the Test Information Function (TIF) is the sum of all item information functions in the test.

RESULT

Table 1 depicts the eigenvalues of the polychoric matrix of math and reading items, it seems that there are 13 underlying factors in the PISA test, rather than 2. Figure 1 displays the loading plot of all item vectors in a 2-factor solution. Obviously, all vectors are jammed together and no clustering pattern can be detected. When math items are analyzed in SAS's MIRT, 7 latent factors are suggested.

Eigenvalues of the Polychoric Correlation Matrix

Eigenvalue	Difference	Proportion	Cumulative
17.1087382	14.6489776	0.3802	0.3802
2.4597606	0.4327083	0.0547	0.4349
2.0270523	0.1973397	0.0450	0.4799
1.8297126	0.2126189	0.0407	0.5206
1.6170937	0.1531681	0.0359	0.5565
1.4639256	0.0526155	0.0325	0.5890
1.4113101	0.1231221	0.0314	0.6204
1.2881880	0.0240435	0.0286	0.6490
1.2641445	0.0471776	0.0281	0.6771
1.2169670	0.0519736	0.0270	0.7042
1.1649934	0.0801460	0.0259	0.7300
1.0848473	0.0717188	0.0241	0.7541
1.0131285	0.0410175	0.0225	0.7767

Table 1

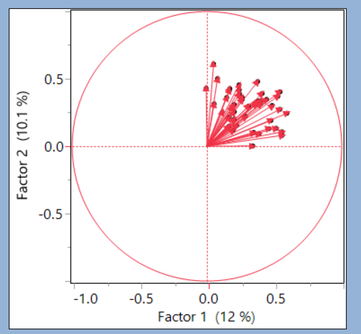


Figure 1

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Does factor indeterminacy matter in multi-dimensional item response theory?

Chong Ho Yu, Ph.D., D. Phil., cyu@apu.edu

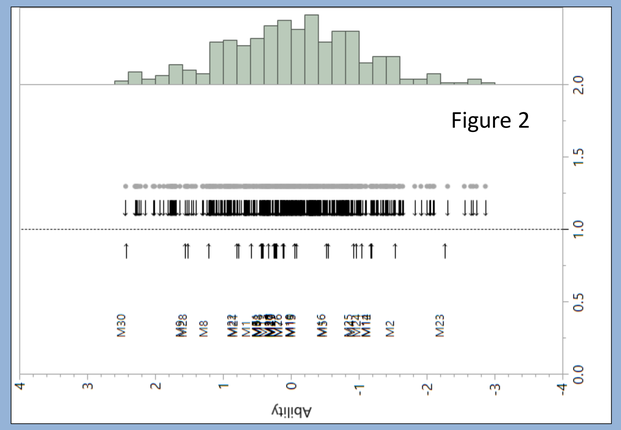
Azusa Pacific University, USA

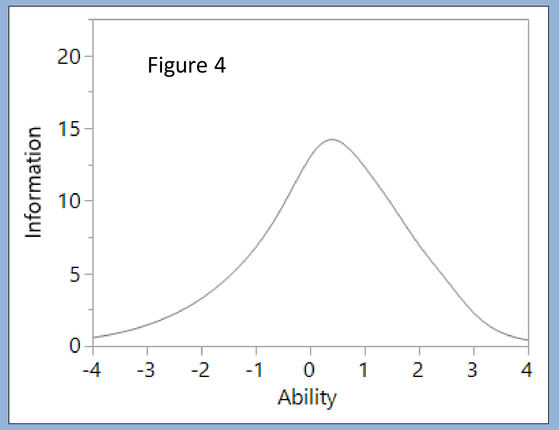
RESULTS CONTINUED

- At first glance these results are puzzling. However, when the dual plot (Figure 2) yielded from a 2–P IRT analysis is examined, it is evident that the math test items are well-written.
  - The item difficulty level is evenly distributed; there is no extremely difficult or super-easy item.
  - The student ability distribution also forms a fairly normal curve.
  - When the item and student attributes compared, it is obvious that the test difficulty matches the student ability.
- The item difficulty/discrimination plot (Figure 3) provides additional support for the psychometric soundness of the test by showing that all items have positive discrimination parameters.
- Test information function plot (Figure 4) shows that much information about users whose ability estimates concentrate around the center (zero) can be learned from the exam. The peak of the TIF is above zero, indicating that more information about the students with slightly above-average ability estimates can be obtained.
- Similar findings are observed in the IRT modeling of reading items.
- Given all these IRT psychometric attributes, it is absurd to deny the validity of the PISA exam just because there is no clear factor solution. Further, when math and reading tests are separately evaluated in two IRT models it returns a single ability estimate for each student. But if MIRT model is used, then it will yield individual ability profiles as test results rather than single scores (Hartig & Hoher, 2009). The question is: could we obtain more useful information by adding this extra layer of complexity?

Item	Difficulty	Discrimination
M1	0.5907	1.4874
M2	-1.5340	0.7029
M3	0.4321	1.5175
M4	-0.9551	1.5605
M5	-0.5497	0.8897
M6	0.4432	1.6813
M7	0.1058	0.6833
M8	1.2082	1.7824
M9	1.5620	2.6163
M10	-0.0576	1.9344
M11	-0.0542	1.2755
M12	-1.1925	1.0609
M13	0.3326	2.4837
M14	-1.1764	0.7011
M15	-0.0782	1.4960
M16	-0.5148	1.2578
M17	0.2201	0.9056
M18	0.4081	1.1894
M19	0.2093	1.5697
M20	0.2474	0.8379
M21	0.7750	2.3186
M22	0.7953	2.1087
M23	-2.2685	1.3003
M24	-1.0449	1.2897
M25	-0.9222	1.4980
M26	0.1243	1.3750
M27	0.2359	0.8302
M28	1.5180	2.2114
M29	0.2076	2.6948
M30	2.4211	2.7404
M31	0.4420	1.6899

Figure 3





CONCLUSIONS

In the past, CTT and IRT were believed to be incompatible. As such, merging these methodologies seemed to be a remote dream. Nevertheless, with the advance of MIRT the dream became a reality. Indeed, combining these methodologies can remediate limitations in both camps. While traditional IRT allows for measurement of a single construct, infusing factor modeling enhances IRT with multidimensionality. In CTT the psychometric property is tied to the entire instrument; consequently, when a researcher wants to customize a scale to a specific population, he or she needs to repeat the tedious process of EFA and CFA. IRT provides users with the flexibility of generating an ad hoc, on-the-fly test. Despite the versatility of MIRT, it is not uncommon that PROC IRT could not yield a sound IRT model and a factor model at the same time. When factor indeterminacy is present, it is advisable to evaluate its psychometric soundness based on IRT alone because content validity can supersede construct validity.

REFERENCES

Gardner, H. (2006). *Multiple intelligences: New horizons in theory and practice*. New York, NY: Basic Book.

Hartig, J., & Hohler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation, 35*, 57-63.

Organization for Economic Co-operation and Development [OECD] (2013). The PISA international database. Retrieved from http://pisa2012.acer.edu.au/

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.