

Effects of TIMSS Sampling Weights on Inference Accuracy when Utilizing Structural Equations Models*

Chih-Chien Yang, PhD
PI, Cognitive Neurometrics Laboratory
Graduate School of Educational Measurement & Statistics
National Taichung University
Noahyang@ntcu.edu.tw

*Parts of the study are financially supported by the National Science Council, Taiwan and some of the current manuscripts are to be incorporated with reports/proposals of the project. Please do not cite without permissions.

Abstract

The project aims to address the issues and promote proper analyses and usages of the TIMSS datasets with appropriate weighting procedures. Literature (e.g., Lohr, 1999; Kish, 1965; Asparouhov, 2005; Yang, & Tsai, 2006) suggested that unequal probability of selection is an inevitable feature of complex sampling surveys as in all TIMSS survey procedures. Not accounting for the impact of complex sampling designs can lead to an underestimate of the sampling variance associated with an estimate and, at the same time, bias the standard error (Yang, 2007). Empirical SEM models (Tsai, & Yang, 2007; Tsai, & Yang, 2007; Yang, & Tsai, 2006) are demonstrated by analyzing TIMSS datasets with sampling weights.

Introduction

TIMSS (Trend of International Mathematics and Science Study) datasets have complex sampling designs (e.g., Gonzalez, & Smith, 1997; Gonzalez, & Miles, 2001); therefore, corresponding proper weighting procedures have to be engaged in any levels of statistical analyses to establish correct inferences. For example, the TIMSS questionnaires were finished by more than 500,000 students and thousands of thousands of parents, teachers and principals. Assisted by more than forty countries scattered around the world, the sample datasets were to infer trends of mathematics and science achievements in these international/gigantic populations. Without proper weighting procedures, biased statistical analyses on the samples can lead misleading conclusions and/or inferences of populations.

The significances of complex sampling procedures are that survey researchers may need to use a sample of subjects sized from a few hundreds to a couple of thousands to infer a population that may be 10 times much larger than the sample. An interesting question may be raised “whether a sample of 1,000 out of a 300,000 population (e.g., the entire elementary students of a grade in Taiwan) is more “representative” than a sample of 100 out of a 3,000 population of interest?” A layman’s answer may be the earlier sample because the number of 1,000 is 10 times greater than 100. While sizes do matter a lot, it is the sampling procedure that is also counted critically. For example, if all of the 1,000 samples are drawn from a specific group of the population, the biased sample may have fewer chances to be representative to the population than a smaller sample but with a well-planned sampling design.

Methods

Statistical literature (e.g., Lohr, 1999; Kish, 1965; Asparouhov, 2005) suggested that unequal probability of selection is an inevitable feature of complex sampling surveys, for example, as those in all TIMSS surveys. In addition, it was proved (Asparouhov, 2005; Yang & Tsai, 2006) that if the unequal probability of selection is not incorporated in the analysis, a substantial bias in the parameter estimates may arise. This bias is called as selection bias. However, if the probability of selection is known and incorporated in the analysis, the selection bias can be eliminated or reduced (Asparouhov, 2005; Yang & Tsai, 2006; Tsai, & Yang, 2007). Analyses performed for complex sampling procedures, therefore, have to incorporate with designated sampling weights to adjust selection biases.

Specifically, Kish (1965) and Kaplan and Ferguson (1999) showed if a sample size n is proportional to its original population size N as in the following equation,

$$\frac{\binom{N-1}{n-1}}{\binom{N}{n}} = n/N,$$

the n/N will then be equal the sampling probability of each n . Assume that n_i denotes the sample size of the i th sampling unit and according to Kish and Kaplan and Ferguson, the probability that the i th sampling unit is selected will be $p_i = n_i/N_i$. Then, a linear weight for each sampling unit can be

$$w_i = \frac{1}{p_i}$$

(Kish, 1965; Kaplan & Ferguson, 1999).

Not only traditional statistical methods, but advanced statistical methods, for example, item response theory (IRT), scaling methods, and structural question models (SEM) also have to be weighted properly to have correct statistical inference for TIMSS study. The project will establish weighting methods that can be incorporated with estimating procedures of these statistical modeling methods. TIMSS datasets of surveys have had geographically-based weights by their original sampling designs for each TIMSS participating country. Yet, dynamic weights not only accounting for geographical strata but also for potentially influential factors, e.g., ethnic groupings, variations in opportunity to learn (OTL), educational stratifications, etc., are needed to achieve proper statistical inferences.

Linear weights that are established by stratified proportions of samples to population (N/n) are well known techniques. Non-linear weights, for example, exponential weights (e.g., Grilli, & Pratesi, 2004) can serve for alternative outcome variables. The project is aimed to study the potential effects of alternative non-linear weights.

Results and Conclusions

Importance of proper usages of these large-scale surveys can not be over-emphasized, yet, some critical issues, e.g., proper TIMSS sampling weights, were often neglected even in published academic papers. Empirical SEM models (Tsai, & Yang, 2007; Tsai, & Yang, 2007; Yang, & Tsai, 2006) are demonstrated by analyzing TIMSS datasets. The project aims to address the issues and promote proper analyses and usages of the TIMSS datasets.

References

- Asparouhov, T. (2005). Sampling Weights in Latent Variable Modeling. *Structural Equation Modeling*, 12, 411-434.
- Gonzalez, E. J., & Smith, T. A. (1997, eds.). *User Guide for the TIMSS International Database*. Chestnut Hill, MA: Boston College.
- Gonzalez, E.J., & Miles, J.A. (2001, eds.). *TIMSS 1999 user guide for the International database: IEA's Repeat of the Third International Mathematics and*

Science Study at the Eighth Grade. Lynch School of Education, Boston College.

Grilli, L., & Pratesi, M (2004). Weighted Estimation in Multilevel Ordinal and Binary Models in the Presence of Informative Sampling Designs. *Survey methodology*, 30, 4-14.

Kaplan, D., & Ferguson, A. J. (1999). On the utilization of sample weights in latent variable models. *Structural Equation Modeling*, 6, 305-321.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.

Lohr, S.L. (1999). *Sampling: Design and Analysis*. Pacific Grove, CA: Duxbury Press.

Tsai, L.T., & Yang, C.C. (2007, accepted). A simulation study on using MIMIC model to assess the accuracy of differential item functioning (in Chinese), *Psychological Testing* (TSSCI), forthcoming.

Tsai, L.T., & Yang, C.C. (2007, accepted). Influence of Sampling Weights on SEM Parameter Estimation Accuracy (in Chinese), *Journal of Education and Psychology* (TSSCI).

Yang, C. C., & Tsai, L. T. (2006). A simulation Study on Computation and Inference Accuracy of Factor Loadings for Large Data Mines. *WSEAS Trans on Information Science and Application* (EI), 12(3), 2577-2579.