Examining Item Parameter Drift as a Source of Construct Shift

Okan Bulut, Luke Stanke, Michael C. Rodriguez,
José R. Palma, Kory Vue, & Julio C. Cabrera
University of Minnesota

Minnesota Youth Development Research Group

April, 2013

Paper presented at the annual meeting of the
American Educational Research Association
San Francisco, CA.

Citation:

Bulut, O., Stanke, L., Rodriguez, M.C., Palma, J.R., Vue, Y.F., & Cabrera, J.C. (2013, April). *Examining item parameter drift as a source of construct shift*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Examining Item Parameter Drift as a Source of Construct Shift

Abstract

Item parameter drift (IPD) is defined as differential change in item parameters or ability estimates over time or testing occasions. Unlike previous IPD studies that have focused on change in item parameters over time, this study examined IPD to detect change in item parameters (construct shift) between grade levels within a single time point. Data for this study come from the Minnesota Student Survey (MSS). Hierarchical general linear modeling (HGLM) framework was used for examining IPD in the items of “being bullied” and “school climate” scales from the MSS. Results indicated that almost all of the items drift across grade levels (6th, 9th, and 12th). This suggests that being bullied and school climate constructs are not invariant across grade levels.

INTRODUCTION

Several assumptions are made about item response theory (IRT) models, including that item parameters are invariant over examinee groups or occasions. Sometimes, however, item parameters may not remain invariant due to factors other than sampling error. When invariance does not hold, the item is considered to be drifting from its original parameter value (Wells, Subkoviak, & Serlin, 2002). To investigate if examinees of the same ability but of different group membership differ in performance on any item, differential item functioning (DIF) methods are applied. In addition to DIF, researchers can be interested in investigating change in a construct over time for the same group of people. When DIF is examined by groups categorized by testing occasions or time-related variables, it is referred to item parameter drift

(IPD; Goldstein, 1983). IPD is defined as differential change in item parameter values or ability estimates over time (Goldstein, 1983).

Investigation of IPD is particularly an important task because if item parameters change over time, it may lead to (1) bias in item parameter estimates, (2) bias in person parameter estimates, and (3) instability in a measurement scale (Babcock & Albano, 2011). Many previous studies have examined the impact of IPD have focused on change in item parameters or theta estimates over time (Babcock & Albano, 2011; Giordano, Subhiyah, & Hess, 2005; Wells, Subkoviak, & Serlin, 2002; Witt, Stahl, Bergstrom, & Muckle, 2003). With psychological scales or measurement tools, IPD may occur among different age or grade levels within the same time point. This means that the same construct can be perceived differently by various age groups, and that using the same survey or scale items across different age groups may result in a construct shift across age or grade levels. For example, the way that a 6th grade student perceives a *being bullied* is different is different from the way a 9th grader would perceive *being bullied*, therefore, each grade should be measured differently.

Minnesota Student Survey (MSS) is such a measurement tool that construct shift is likely to occur. MSS is designed and administered by an interagency team from the Minnesota Departments of Education, Health and Human Services, Public Safety, and Corrections. The survey aims to monitor important trends and support planning efforts of local public school districts and the four collaborating state agencies. The survey is administered every three years to students in 6th, 9th, and 12th grades. It includes different types of survey items that are asking students' opinions about risky behaviors and school environment. From the items in MSS, several subscales were defined including: school climate, being bullied, and community support.

In the MSS, the relative direction of a construct shift is associated with the content of the survey items, and reflects changing sensitivity against risky behaviors and issues related to school environment. Because older students may already have been exposed to risky behaviors or negative school environment for a long time, they may be less disgruntled about these issues. On the contrary, young students can be more sensitive against these issues, and therefore respond survey items about these issues differently. This situation may result in a construct shift in the MSS scale across different age (or grade) levels. This study examines to what extent grade level (i.e. time) affect the way students response survey items, and change the construct intended to be measured. IPD analysis was used for detecting items that tend to shift across age levels.

METHOD

Data Source

Data for this study come from the 2010 Minnesota Student Survey. In 2010, 130,908 students participated in the MSS. For this study, a random sample of 40,000 students was taken from the whole sample. Two scales were chosen from the MSS to examine IPD (i.e. construct shift). These scales were "*being bullied*" and "*school climate*". Table 1 includes sample sizes for each grade for the 2010 MSS data. Percentages across grades are for each year cohort. Gender was split evenly between females and males, and the mean age at each year was 14 years.

Insert Table 1 here

Analysis

To analyze the impact of IPD in the MSS item, hierarchical generalized linear modeling (HGLM) was used. The HGLM framework can accommodate nested data structures while incorporating additional covariates at the item, person, and other grouping levels (e.g., Kamata, 2001; Pastor, 2003). In each model, persons are clusters, items are the repeated observations, and binary responses are the dependent variable. Responses are denoted as Y_{ij} = 0 or 1, with j = 1,…., J as an index for persons, and i = 1, ……, I as an index for items. Y_{ij} has a Bernoulli distribution with π_{ij} which is the expected value of Y_{ij} based on the probability. There is a link function which puts π_{ij} into a continuous scale between -∞ and +∞. In the context of IRT, this link function is the logit function:

$$\eta_{ij} = ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) \tag{1}$$

By using this logit function, the traditional Rasch model (Rasch, 1960) can be written in the HGLM format for N items as:

$$\begin{aligned} \eta_{ij} &= \beta_{0j} + \sum_{q=1}^{N-1} \beta_{qj} X_{qij} \\ \beta_{0j} &= \gamma_{00} + \mu_{0j} \\ \beta_{1j} &= \gamma_{10} \\ &\ \vdots \\ \beta_{(N-1)j} &= \gamma_{(N-1)0} \end{aligned} \tag{2}$$

where the intercept $\beta_{0j}$is the parameter for a the reference item, and the other terms (β_{qj}) are the parameters for the remaining items as differences from the reference item. X_{qij} is the item indicator variable (X_{qij} = 1 when q = i, or X_{qij} = 0 otherwise) (Kamata, 2001). Using the same framework, the IPD model can be defined by adding a person covariate (W_j) for the age level:

$$\eta_{ij} = \beta_{0j} + \sum_{q=1}^{N-1} \beta_{qj} X_{qij}$$

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} W_{0j} + \mu_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} W_{1j} \\ &\vdots \\ \beta_{(N-1)j} &= \gamma_{(N-1)0} + \gamma_{(N-1)1} W_{(N-1)j} \end{aligned} \tag{3}$$

In Equation 3, the term γ_{q1}represents the expected change in item location β_{qj} on the logit metric for a one unit change in the age level (W_j).

In this study, expected change (i.e. construct shift) in item location due to time was defined as a linear IPD or a categorical IPD (i.e. factor IPD). Grade level was used as the time variable in all models. In the linear IPD model, grade level was used a continuous person-level covariate whereas in the factor IPD model, grade level was used a categorical predictor representing age groups in the model. To examine the impact of IPD on the MSS items in *being bullied* and *school climate* scales, all items were first analyzed concurrently ignoring the effect of age. Then, linear and factor IPD models were estimated using the same items. Concurrent, linear IPD and factor IPD models were compared in terms of model-fit to investigate the overall effect of the construct shift. Also, all items were examined individually for IPD in linear and factor IPD models. Lastly, the correlation between scale scores from the three models and self-reported GPA was obtained to show the impact of construct shift across grade levels. Concurrent, linear IPD and factor IPD models were estimated via the *lme4* package (Bates, Maechler, & Bolker, 2012) in R. Since the *lme4* package requires dichotomous data, all items in *being bullied* and *school climate* scales were dichotomized.

RESULTS

Table 2 contains the fit statistics for the three models' fit to the data. Based on the AIC and BIC, the best fitting model is the Factor IPD model. However, the Linear IPD model also fit significantly better than the Concurrent Model. Both the Linear IPD and Factor IPD models suggest a changing of the construct as students improve from grade to grade.

Insert Table 2 here

Table 3 contains the results for the concurrent, linear IPD, and factor IPD models. All of the covariates for grade in the linear IPD and factor IPD models were statistically different than zero at a *p*-value of at least .05. The coefficients for the concurrent Model represent the locations of items for a person, regardless of grade. For the linear IPD Model, the first coefficient for each item represents the location of endorsing the item for 6th grade students, while the second coefficient represents a change in item location depending on grade, this could also be considered the slope for difficulty change of the item parameters. For the Factor IPD Model, the first parameter is location the item for 6th graders, the second parameter is the location of the item for 9th graders, and the third parameter is the location of the item for 12th graders.

Insert Table 3 here

Figure 1 displays the plots item locations for each the eight items used to measure *being bullied*. The x-axis consists of grade year and the y-axis is the location of the item in terms of logits. The dashed line represents the location of the item for the concurrent model. The solid line represents the item locations for the linear IPD model, and the points on each graph represent the item locations for the factor IPD model.

Six of the eight items seem to show linear progression of the item locations, as the location of the items using the Factor IPD model are within two standard error intervals of the estimates of item locations using the Linear IPD model. From the figure, it can be seen that each item location changes based on grade. This suggests that the construct of being bullied is different for 6^{th}, 9^{th}, and 12^{th} graders.

Table 4 displays a table of unstandardized correlations between the *being bullied* scale and the self-reported GPA of students. The correlations are organized by method used to calibrate the *being bullied* scale and by subset of the sample used for this analysis. While correlations are similar for all three models, the Linear and Factor Models have the closest relationship. All correlations suggest that the more a student is bullied, the lower a GPA a student is likely to have. Table 5 displays the correlation of the person estimates from the EAP. Most notably, all of the scales are highly correlated; however the factor IPD and linear IPD models have a correlation of .999, suggesting that the models are nearly identical. The preliminary results in this paper didn’t include school climate scale. The same IPD analyses will be conducted for this scale, and the results will be reported in the final paper.

DISCUSSION

This study demonstrates a different application of item parameter drift (IPD) analysis using the HGLM framework. Unlike most IPD studies focusing on the change in item location over time, this study used IPD for detecting construct shift across different grade levels within the same time point. The IPD model provides estimates of the effect of IPD on each item parameter simultaneously controlling for person ability. Within the HLGM framework, both linear and factor IPD models can be run. Results of this study indicated that items measuring psychological constructs may show drift across different grade levels, which may result in construct shift. Therefore, when items are expected to function differently across grade levels, the target construct should be estimated for each grade level separately.

References

Babcock, B., & Albano, A. D. (2011). *Rasch scale drift over time: Examining when to reset the scale.* Paper presented at the meeting of the National Council on Measurement in Education, New Orleans, LA.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4.

Giordano, C., Subhiyah, R., & Hess, B. (2005, April). *An analysis of item exposure and item parameter drift on a take-home recertification exam.* Paper presented at the annual meeting of the American Educational Research Association , Montreal, Quebec, Canada.

Goldstein, H. (1983). Measuring changes in educational attainment over time: Problems and possibilities. *Journal of Educational Measurement, 20,* 369-377.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement, 38*, 79 – 93.

Rasch G (1960*). Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research, Copenhagen.

Wells, C. S., Subkoviak, M. J., & Serlin, R. C. (2002). The effect of item parameter drift on examine ability estimates. *Applied Psychological Measurement, 26*, 77-87.

Witt, E.A., Stahl, J.A., Bergstrom, B.A., & Muckle, T. (2003, April). *Impact of item drift with non-normal distributions*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Table 1

A breakdown of sample size by count and proportion across grade.

Grade	N	%
6th	14,136	.353
9th	14,362	.366
12th	11,222	.281
Total	40,000	1.000

Table 2

Fit statistics for three IPD models

	Parameters Estimated	AIC	BIC	LL	Deviance
Concurrent	9	238554	238649	-119268	238536
Linear DIF	17	231498	231678	-115732	231464
Factor DIF	23	230780	231023	-115367	230734

Table 3

Parameter estimates and standard errors for the three models fit to the data.

Coefficient	Concurrent	Linear DIF	Factor DIF
G00	-1.28 (0.01)**	-0.95 (0.02)**	-0.97 (0.02)**
G01	---	-0.13 (0.01)**	-1.32 (0.02)**
G02	---	---	-1.73 (0.03)**
G10	-2.12 (0.01)**	-1.64 (0.02)**	-1.67 (0.02)**
G11	---	-0.20 (0.01)**	-2.21 (0.02)**
G12	---	---	-2.84 (0.03)**
G20	-0.75 (0.01)**	0.08 (0.02)**	0.07 (0.02)**
G21	---	-0.31 (0.01)**	-0.81 (0.02)**
G22	---	---	-1.86 (0.03)**
G30	-1.84 (0.01)**	-1.00 (0.02)**	-1.01 (0.02)**
G31	---	-0.35 (0.01)**	-2.03 (0.02)**
G32	---	---	-3.09 (0.04)**
G40	-5.45 (0.04)**	-5.67 (0.08)**	-5.89 (0.09)**
G41	---	0.05 (0.02)*	-5.30 (0.07)**
G42	---	---	-5.56 (0.08)**
G50	-2.01 (0.02)**	-1.94 (0.06)**	-2.06 (0.02)**
G51	---	-0.06 (0.01)**	-2.39 (0.03)**
G52	---	---	-1.43 (0.02)**
G60	-1.45 (0.01)**	-1.16 (0.06)**	-1.92 (0.03)**
G61	---	-0.12 (0.01)**	-5.26 (0.07)**
G62	---	---	-2.62 (0.03)**
G70	-3.07 (0.02)**	-4.34 (0.01)**	-2.43 (0.03)**
G71	---	-0.37 (0.01)**	-0.97 (0.02)**
G72	---	---	-1.32 (0.02)**

*Note: * = .05 p-value, ** < .01 p-value*

Table 4

A breakdown of unstandardized correlations between the being bullied scale and GPA by sub-sample and scaling method.

	Concurrent	Linear	Factor
6th Grade	-0.158	-0.161	-0.162
9th Grade	-0.226	-0.226	-0.225
12th Grade	-0.206	-0.203	-0.203
Overall	-0.191	-0.202	-0.199

Table 5

Inter-correlation of person estimates by estimation method

	Conc.	Line.	Fact.
Concurrent	1.000	---	---
Linear	0.980	1.000	---
Factor	0.979	0.999	1.000

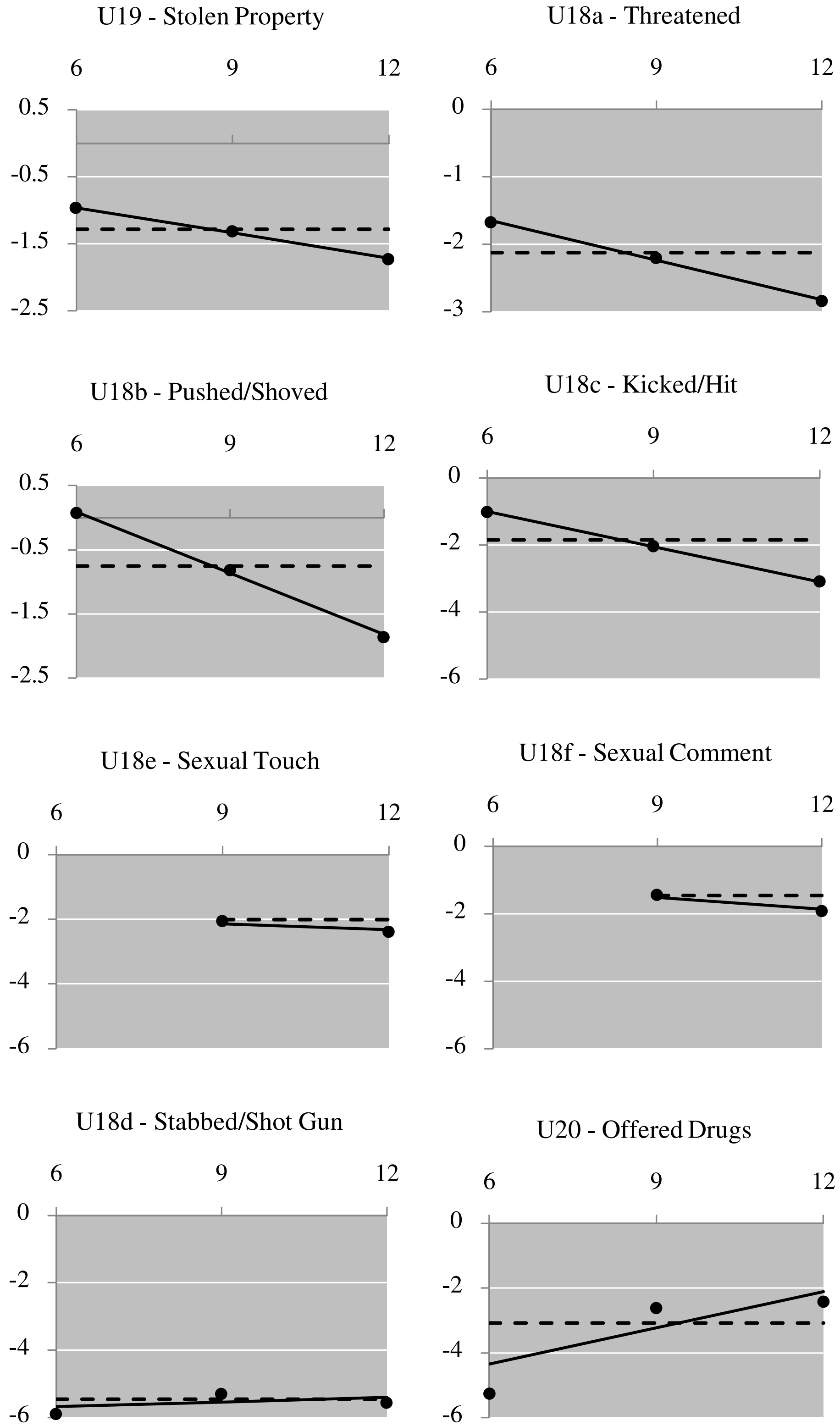

Figure 1. Plots of item locations for each of the 8 items.