Investigating and Comparing the Item Parameter Drift in the Mathematics Anchor/Trend Items in TIMSS between Singapore and the United States

Amery D. Wu
Zhen Li
Siok Leng Ng
Bruno D. Zumbo

Abstract

This paper studied the item parameter drift (IPD), uniform and non-uniform, of the trend items in the Trends in International Mathematics and Science Study (TIMSS). TIMSS has maintained and used an item pool from which items have been repeatedly administered since 1995. A problem with long-term usage of test items is that over time the items might perform differently as a result of change in measurement precision and/or difficulty level. This change is referred to item parameter drift. In this paper, the 23 trend items of Grade-8 mathematics test administered across three TIMSS cycles were examined using logistic regression analysis to see whether IPD was present. In addition, the IPD results were compared between Singapore and the United States. The results showed that, for both countries, neither uniform nor non-uniform IPD had occurred.

Introduction

For large-scale achievement assessment, a set of items is often maintained and secured for repeated use. These repeatedly administered items typically function as the anchor items for linking and equating purposes or as trend items for investigating changes in performance over time. The premise and justification for the repeated use of such anchor/trend items are that the items perform identically for the target population across repeated use. That is, the precision (i.e., discrimination power) and the difficulty level of the anchor/trend items remain stable over repeated administrations. A violation of such a premise is referred to as *item parameter drift* (IPD) (Goldstein, 1983).

IPD poses a threat to measurement applications that require a stable scale (Well, Subkoviak, & Serlin, 2002). For example, based on the belief that the anchor items perform the same across administrations, an examinee's true ability is estimated in order to make a decision about the individuals' admission to graduate studies (e.g., GRE) (Miller, Gesn, & Rotou, 2005). If the anchor items have displayed IPD, linking or equating based on such drifted anchor items may result in incomparable scores across administrations and lead to biased and inequitable decisions.

In practice, IPD is likely to occur over prolonged use of test items and result in the anchor/trend items becoming relatively less precise or difficult. IPD may occur due to a change in curriculum coverage or increase in teaching and exercise (Bock, Muraki, & Pfeiffenberger, 1988; Goldstein, 1983; Mislevey, 1982). Immense teaching-to-test and test preparation could change how an item originally performs. Test items may also display IPD because of excessive item exposure or poor control of security.

Up to date, IPD has not received much attention empirically and only a few studies have conducted IPD analyses in the educational and psychological measurement literature (Juve, 2004; Pleysier, Pauwels, Vervaeke, & Goethals, 2005; Skykes & Ito, 1993). Specifically, no empirical study has investigated or publicized the existence of IPD for Trends in International Mathematics and Science Study (TIMSS), the largest and most complex international study of students' achievement ever conducted. Three cycles of TIMSS, 1995, 1999, and 2003, have been conducted since it was first introduced. Some of the items in the item pool have been used repeatedly over the three cycles. These anchor/trend items were believed to be absent from IPD and were qualified to provide a common metric for linking and equating purposes or study performance trends over time. However, no study has tested the assumption that IPD was not present in the TIMSS anchor/trend items.

Over the years since TIMSS was first introduced in 1995, it is reasonable to conjecture that some of these anchor/trend items may have displayed IPD due to some of the aforementioned reasons. Also, it is reasonable to conjecture that the prevalence of IPD may vary across TIMSS participating countries because of different educational culture and practices. For example, the custom of teaching-to-test, a potential reason for IPD, is immensely prevalent in the top performing countries such as Singapore, Korea, Hong Kong, Taiwan, and Japan (Dolly, 1992; Johnson & Johnson, 1996; Pettersen, 1993; Sharma, 1997). If IPD exists in some of the anchor/trend items of TIMSS, scores linked and equated across countries and cycles based on such items might be incomparable and lead to meaningless and invalid interpretations.

The first purpose of this study was to examine whether IPD occurred in the anchor/trend items of the TIMSS Grade-8 mathematics test over the three administration cycles. More specifically, the following research questions were addressed: (1) Have the anchor/trend items in TIMSS drifted from 1995 to 1999 (cycle-1 to cycle-2)? (2) Have the anchor/trend items in TIMSS drifted from 1995 to 2003 (cycle-1 to cycle-3)? The second purpose was to compare the prevalence in, and pattern of IPD, if IPD was shown to be present, between the United States and Singapore.

At this point, it is important to point out that the notion of IPD is conceptually parallel to that of a widely used statistical technique to test item bias, *differential item functioning* (DIF). The only difference is that DIF examines whether items function differently between groups that are defined on differences in examinees' individual characteristics such as gender, ethnic group, or country whereas the group membership in a IPD study is defined on the time when examinees take a test (Rupp & Zumbo, 2006). Ultimately, the foundations of IPD and DIF are both rooted in the notion of investigating *measurement invariance*.

Meredith (1993) defined measurement invariance as the examinee's probability of an item score does not depend on the examinee's group membership, given an examinee's true score. Over the years, a variety of statistical frameworks such as item response theory, Mantel-Haenszel, and ordinal logistic regression have been developed to empirically test the presence of measurement invariance at the item level (see review in Camilli & Shepard, 1994; Clauser & Mazor, 1998; Donoghue & Isham, 1998). In this paper, logistic regression (LogR) was chosen to address our three research questions. A more detailed description and justification for choosing the LogR IPD method was presented in the methodology section.

Methodology

Instrument and Data

Data from TIMSS 1995, 1999, and 2003 Grade-8 mathematics tests were chosen to examine whether the anchor/trend items had displayed IPD. Data were collected in each country using clearly specified standardized procedures at approximately the same time of the academic year. There were 23 anchor/trend items in the TIMSS Grade-8 item pool over the three cycles. The 23 anchor/trend items were all scored with two scoring points (incorrect, 0 or correct, 1). The items included five major content areas: (1) Fractions and Number Sense; (2) Measurement; (3) Geometry; (4) Algebra; and (5) Data Representation, Analysis and Probability.

Because of the matrix sample design of the items (Martin & Kelly, 1998), examinees across three cycles did not answer the same 23 anchor/trend items. In order to calculate a total score based on the same number of items, i.e., the same maximum total score for each examinee, which was needed for the LogR IPD method, missing data were imputed using maximum likelihood method in PRELIS (Jöreskog & Sörbom, 1999). Missing data imputation was justifiable for this study because the data-missing pattern was considered missing completely at random (i.e., MCAR, Pigott, 2001; Rubin, 1976) as a result of random assignment of the booklets (hence, items) to the examinees.

Participants

Data from students in Singapore and the United States were selected for comparison purpose. All the participants across three cycles were grade eight students and were around 13 years old at the time of testing. Sample size for Singapore was 15,248 (cycle-1: 8,285, cycle-2: 4,966, and cycle-3: 1,997); for the United States was 22,958 (cycle-1: 10,948, cycle-2: 9,034, and cycle-3: 2,976). Gender composition was close to equal for both countries.

IPD Statistical Analyses

In the present study we adopted the LogR framework to detect IPD. There were three major reasons for using LogR. Along with the description of the three reasons, we weaved our discussion on how LogR analyses can be used to detect IPD in the next paragraphs.

The first reason for choosing LogR for our IPD study was that most of the IPD/DIF statistical methods such as IRT or Mantel-Haenszel cannot simultaneously detect IPD for more than two groups (DeMars, 2004; Kim, Cohen, & Park, 1995). Taking our research questions for example, there were three groups involved in the study, cycle-1, cycle-2, and cycle-3 and our primary interest was only to compare whether items have displayed IPD from cycle-1 to cycle-2 and from cycle-1 to cycle-3. For IPD methods other than LogR, if more than two groups were involved, analyses must be conducted on a paired basis for each item: one detecting IPD between cycle-1 and cycle-2 and the other between cycle-1 and cycle-3.

A great advantage of using LogR is that the grouping variable, like multiple regression, is able to include as many groups as needed. In addition, the researchers can specify the planned comparisons of their interest by using various contrast doing schemes provided in most of the popular statistical softwares. For the present study, the two comparisons of interest were specified by "indicator contrasting" using cycle-1 as the reference group in SPSS (see Table 1). For the three cycles, only two coding variables, in contrast to two statistical analyses, are needed. Coding variable-1 contrasted the IPD investigation between cycle-1 and cycle-2 and coding variable-2 contrasted the IPD investigation between cycle-1 and cycle-3. Hence, for each item, the two IPD investigations of our interest were analyzed in one single LogR analysis by including all three groups in the grouping variable, "cycle".

Table 1. Contrast (Indicator) Coding for "Cycle"

	Parameter Coding	
	Variable-1 Cycle-1 vs. Cycle-2	Variable-2 Cycle-1 vs. Cycle-3
Cycle-1	0	0
Cycle-2	1	0
Cycle-3	0	1

The second reason for using LogR was implied in the definition of IPD. Following our earlier introduction on IPD, a statistical IPD analysis can be defined as a procedure that matches examinees on the true score to see if comparable examinees tested at different cycles performed the same on the anchor/trend items. In words, statistical IPD analysis aims to detect whether examinees from different cycles, who are matched on the ability, have the same probability of getting an item right. This definition implies that IPD analysis can be viewed as a model-based sequential regression analysis where the item score is the response variable, total score, a proxy of the true score[1], functions as a matching variable (i.e., covariate), and the variable "cycle" and the interaction variable "total by cycle" are the explanatory variables. Because the 23 anchor/trend items were binary items scored as 0 or 1, the theoretically appropriate and commonly practiced regression analysis is, naturally, LogR.

This sequential LogR IPD analysis entailed three explanatory variables that entered the regression model in the following sequence,

Model-1: Logit = $b_0 + b_1$*Total (df = 1)
Model-2: Logit = $b_0 + b_1$*Total + b2*Cycle (df = 1+2= 3)
Model-3: Logit = $b_0 + b_1$*Total + b2*Cycle + b3*Total by Cycle (df = 1+2+2= 5)

Note that the dependent variable was written in the logit form, the natural logarithm of the probability of getting an item right, so that the relationship was linear in the coefficients (see Hosmer & Lemeshow, 2000 for details).

In model-1, the "total score" entered the model with one degree of freedom (df) and functioned like a matching variable. In model-2, the "cycle" variable also entered the model to detect uniform IPD. The omnibus test of model coefficients (coding variable-1 and coding

[1] Empirical evidence that the data structure among the 23 items was unidimensional was necessary if the item scores were to add up to a total score to proxy the examinees' true score. Using factor analysis and parallel analysis to decide on the dimension (i.e., number of factors), the 23 items were shown to be unidimensional for both Singapore and the United States students, thus, the use of total score was justified.

variable-2) was examined by Chi-square statistics with two degrees of freedom. The two degrees of freedom referred to the two contrast coding variables for the three cycles as discussed earlier. An item would be flagged as displaying uniform IPD if the omnibus test of the model coefficients in model-2 was significant with two degrees of freedom. In model-3, the interaction term "total by cycle" also entered the model to detect non-uniform IPD with two degrees of freedom. The two degrees of freedom referred to the two interaction coding variables (i.e., "total by coding variable-1" and "total by coding variable-2"). Non-uniform DIF referred to the effect of the administration cycle on the probability of getting an item right was variant along the examinees' total score continuum. For example, an item might not display IPD for examinees with lower total score but for examinees with higher total score. If such interaction was present, non-uniform IPD was said to be existent.

In sum, LogR IPD analysis tested the null hypotheses that the effects of "cycle" and "total by cycle" were both zero in the population. If either of the omnibus tests of coefficients in model-2 or model-3 was significant, IPD was considered to be statistically present. If IPD was detected, significant tests on individual coefficients using Wald statistics with one degree of freedom would be examined to see where IPD had occurred for uniform IPD (i.e., "cycle-1 vs. cycle-2" and/or "cycle-1 vs. cycle-3") as well as for non-uniform IPD (i.e., "total by coding variable-1" and/or "total by coding variable-2"). Because 23 analyses were involved, the significant alpha level for all hypothesis tests was adjusted at the 0.002 level (0.05/23).

Also, to improve the statistical matching, a "purified total" was calculated for each examinee by subtracting the item score from the total score. Consequently, the purified total was used for all the LogR sequential modelling instead of the original total score. Detailed discussions of sequential based LogR DIF/IPD analysis can be found in Swaminathan and Rogers (1990) and Zumbo (1999). The SPSS syntax for the above described LogR IPD method was given in Appendix A.

The third reason for using LogR was related to the issue of practical significance of hypothesis testing of IPD. In addition to hypothesis testing, typically, effect size measures are incorporated to interpret the magnitude of IPD and determine whether IPD is negligible. Theoretically, statistical hypothesis testing will detect trivial effect if the sample size is large and lose its practical usefulness. For this reason, incorporating effect size measures was necessary for this study because not only a substantially large number of examinees were sampled in the TIMSS assessment to represent the population in a country but also the sample sizes of the three cycles were aggregated into the variable "Cycle" for each country in order to investigate IPD.

On this note, another advantage of adopting LogR for this study was that cut-off values for appropriate interpretation of effect sizes have been proposed and investigated along with LogR IPD analyses (Hidalgo & Lopez-Pina, 2004; Jodoin & Gierl, 2001; Zumbo, 1999). In this study, we employed Jodoin and Gierl's (2001) effect size criteria to quantify the magnitude of IPD. If the change in Nagelkerke R-square between model-1 and model-2 (for uniform IPD) or between model-2 and model-3 (for non-uniform IPD) was less than 0.035, IPD would be considered as negligible, between 0.035 and 0.070 as moderate, or greater than 0.070 as large.

Results

The results for the omnibus test of model coefficients were shown in Table 2. For uniform IPD (model-2), as expected, with large sample size, Chi-square test flagged most of the items as displaying uniform IPD for both Singapore (20 out of 23) and the United States (18 out of 23) using the $p< 0.002$ criterion. However, for both Singapore and the United States, the effect sizes for the IPD items were all considered negligible using the change in Nagelkerke R-square (ΔR^2) < 0.035 criterion. All in all, the 23 anchor/trend items had not displayed non-negligible uniform IPD in both countries, neither from cycle-1 to cycle-2 nor from cycle-1 to cycle-3.

As for non-uniform IPD (model-3), Chi-square flagged eight items as displaying non-uniform IPD for Singapore using the $p< 0.002$ criterion and 20 items for the United States. However, as in the uniform IPD case, the effect sizes for the IPD items were all considered negligible for both Singapore and the United States using the change in Nagelkerke R-square < 0.035 criterion. All in all, the 23 anchor/trend items had not displayed non-negligible non-uniform IPD in both countries, neither from cycle-1 to cycle-2 nor from cycle-1 to cycle-3. Because the omnibus test for neither the uniform IPD nor the non-uniform IPD identified non-negligible IPD using the combination rules of Chi-square test and change in effect sizes, no further tests on individual coefficients using Wald statistics was examined and reported.

Table 2. Omnibus Test of Model Coefficients For Uniform and Non-uniform IPD

Country	Singapore						United States					
Model	Model-2 (Uniform IPD)			Model-3 (Non-uniform IPD)			Model-2 (Uniform IPD)			Model-3 (Non-uniform IPD)		
Item	X^2(2df)	p	ΔR^2	X^2(2df)	p	ΔR^2	X^2(2df)	p	ΔR^2	X^2(2df)	p	ΔR^2
Item1	73.440	**0.000**	0.007	2.858	0.240	0.000	25.920	**0.000**	0.001	96.513	**0.000**	0.003
Item2	83.847	**0.000**	0.009	0.401	0.818	0.000	89.169	**0.000**	0.004	6.930	0.031	0.005
Item3	42.282	**0.000**	0.005	1.817	0.403	0.000	55.700	**0.000**	0.002	94.262	**0.000**	0.004
Item4	69.745	**0.000**	0.006	67.701	**0.000**	0.006	23.585	**0.000**	0.001	298.576	**0.000**	0.013
Item5	179.896	**0.000**	0.015	61.028	**0.000**	0.005	30.114	**0.000**	0.002	230.853	**0.000**	0.012
Item6	15.697	**0.000**	0.003	1.730	0.421	0.000	111.256	**0.000**	0.006	9.617	0.008	0.001
Item7	43.365	**0.000**	0.008	239.000	**0.000**	0.001	10.740	0.005	0.000	23.413	**0.000**	0.001
Item8	51.103	**0.000**	0.008	6.538	0.038	0.000	5.783	0.055	0.000	81.214	**0.000**	0.004
Item9	10.561	0.005	0.001	0.112	0.946	0.000	1.233	0.540	0.000	64.631	**0.000**	0.002
Item10	15.322	**0.000**	0.001	1.441	0.487	0.000	90.633	**0.000**	0.004	8.126	0.017	0.000
Item11	17.854	**0.000**	0.002	14.213	**0.001**	0.001	4.077	0.130	0.000	25.054	**0.000**	0.001
Item12	98.321	**0.000**	0.011	20.388	**0.000**	0.002	83.189	**0.000**	0.005	88.850	**0.000**	0.004
Item13	46.291	**0.000**	0.005	10.616	0.005	0.001	124.517	**0.000**	0.008	107.853	**0.000**	0.007
Item14	18.747	**0.000**	0.002	11.725	0.003	0.001	11.99	0.002	0.001	46.470	**0.000**	0.001

Country	Singapore						United States					
Model	Model-2 (Uniform IPD)			Model-3 (Non-uniform IPD)			Model-2 (Uniform IPD)			Model-3 (Non-uniform IPD)		
Item	X^2(2df)	p	ΔR^2	X^2(2df)	p	ΔR^2	X^2(2df)	p	ΔR^2	X^2(2df)	p	ΔR^2
Item15	48.811	**0.000**	0.009	0.326	0.850	0.000	287.555	**0.000**	0.012	23.911	**0.000**	0.001
Item16	41.352	**0.000**	0.003	19.663	**0.000**	0.001	24.662	**0.000**	0.000	30.817	**0.000**	0.001
Item17	17.887	**0.000**	0.001	15.806	**0.000**	0.002	73.451	**0.000**	0.000	87.888	**0.000**	0.001
Item18	60.861	**0.000**	0.005	23.987	**0.000**	0.002	214.736	**0.000**	0.008	87.888	**0.000**	0.002
Item19	15.998	**0.000**	0.006	1.030	0.598	0.001	74.153	**0.000**	0.004	35.935	**0.000**	0.002
Item20	13.291	**0.000**	0.002	4.070	0.131	0.000	37.172	**0.000**	0.001	35.856	**0.000**	0.002
Item21	6.929	0.031	0.002	2.274	0.321	0.001	18.818	**0.000**	0.001	24.868	**0.000**	0.002
Item22	5.557	0.062	0.002	0.088	0.957	0.000	90.662	**0.000**	0.004	45.015	**0.000**	0.001
Item23	51.860	**0.000**	0.004	8.824	0.012	0.001	759.965	**0.000**	0.003	24.661	**0.000**	0.001

Note. ΔR^2: change in Nagelkerke R-square
Note. $p < 0.002$ was highlighted in bold.

Conclusions and Discussions

The purpose of this paper was to examine whether item parameter drift (IPD) occurred in the anchor/trend items of TIMSS Grade-8 mathematics test over the three administration cycles as well as to compare the prevalence in, and pattern of IPD between Singapore and the United States. Using the Chi-square test accompanied with the effect size rule, no non-negligible IPD was found in any of the 23 anchor/trend items and the results were the same for Singapore and the United States. In other words, this finding provides a piece of evidence that the items have been performing stably for the target population across repeated uses for Singapore and the United States. Hence, past linking and equating exercises and trend research for Singapore and the United States based on these anchor/trend items are assured to be trustworthy.

This study also demonstrated that incorporating effect sizes to the IPD decision, in addition to hypothesis testing, is essential for large-scale assessment like TIMSS. The Chi-square test flagged most of the items as displaying uniform IPD, yet the effect sizes showed that the statistically significant IPD effects were so trivial that IPD for all flagged items could be ignored.

Note that this study only investigated the existence of IPD in the Singapore and the United States data. Our finding does not generalize to data from other countries. Our recommendations to the TIMSS administration is to investigate IPD of the anchor/trend items across all participating countries on a regular basis in order to circumvent the potential biases or inequity problems if the same anchor/trend items are continuously used in the coming administrations.

References

Bock, R. D., Muraki, E., & Pfeiffenberger, W. (1988). Item pool maintenance in the presence of item parameter drift. *Journal of Educational Measurement, 25*, 275-285.

Camilli, G., & Shepard, L. A. (1994). *Method for identifying biased test items* (Vol. 4). Thousand Oaks, CA: Sage Publications.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31-44.

DeMars, C. E. (2004). Detection of item parameter drift over multiple test administrations. *Applied Measurement in Education, 17*, 265-300.

Dolly, J. P. (1992, June). "*Juku" and the performance of Japanese students: An American perspective.* Paper presented at the Annual Meeting of the Japan United States Teacher Education Consortium, Tokyo, Japan.

Donoghue, J. R., & Isham, S. P. (1998). A comparison of procedures to detect item parameter drift. *Applied Psychological Measurement, 22*, 33-51.

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

Hidalgo, M. D., & Lopez-Pina, J. A. (2004). Differential item functioning detecting and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement, 64*, 903-915.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*, 2nd ed. New York: Wiley.

Jodoin, M. G., & Gierl, M. J. (2001). Type-one error and power rates using an effect size measure with the logistic regression for DIF detection. *Applied Measurement in Education, 14,* 329-49.

Johnson, M. L. & Johnson, J. R. (1996). Daily life in Japanese high schools (Report No. EDO-SO-96-7). Washington, DC: Office of Educational Research and Improvement (EDD00036).

Jöreskog, K.G., & Sörbom, D. (1999). User's Reference Guide, Scientific Software International.

Juve, J. A. (2004). Assessing differential item functioning and item parameter drift in the college basic academic subjects examination. *Dissertation Abstracts International Section A: Humanities & Social Science*, *65(6-A),* p.2088.

Kim, S. H., Cohen, A. S., & Park, T. H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement, 32*, 261-276.

Martin, M. O., & Kelly, D. L. (1998). *TIMSS technical report* (Vol.1)*: Design and development.* Chestnut Hill, MA: TIMSS International Study Center.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-43.

Miller, G. E., Gesn, P. R., & Rotou, O. (2005). Expected linking error resulting from item parameter drift among the common items on Rasch calibrated tests. *Journal of Applied Measurement, 6*, 48-56.

Mislevey, R. J. (1982, April). *Five steps toward controlling item parameter drift*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Pettersen, L. (1993). Japan's cram schools. *Educational Leadership, 50*, 56-58.

Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation, 7*, 353-383.

Pleysier, S., Pauwels, L., Vervaeke, G., & Goethals, J. (2005). Temporal invariance in repeated cross-sectional 'Fear of Crime' research. *International Review of Victimology*, *12*, 273-292.

Sharma, Y. (1997). Tough fight for places among tiny elite. *Times Educational Supplement*, 4206, p. 16.

Skykes, R. C., & Ito, K. (1993, April). *Item parameter drift in IRT-based licensure examinations.* Paper presented at the Annual Meeting of the National Council on Measurement in Education. Atlanta, GA.

Rubin, D. B. (1976). *Inference and missing data. Biomerika, 63*, 581-592.

Rupp, A. A., & Zumbo, D. B. (2006). Understanding parameter invariance in unidimensional IRT models. *Educational & Psychological Measurement, 66*, 63-84.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370.

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

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Appendix A

Syntax for Logistic IPD Method

```
---------------------------------------------------------------------------------------------------------------------------------------------
GET
  FILE='your file name'.

* dv is the item being analysed.

compute dv=v1

* ==========================================================.

* compute the purified total score, the total minus the item being analysed.

COMPUTE purified_total = sum(v1 to v23)-dv .
EXECUTE .

LOGISTIC REGRESSION  dv
  /METHOD = ENTER purified_total  /METHOD = ENTER cycle  /METHOD = ENTER purified_total*cycle
  /CONTRAST (cycle)=Indicator(1)
  /SAVE = PRED
  /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .

GRAPH
  /SCATTERPLOT(BIVAR)=purified_total WITH PRE_1 BY cycle
  /MISSING=LISTWISE .
```