A Short Note on Estimating the Testlet Model With Different Estimators in Mplus

Educational and Psychological
Measurement

2018, Vol. 78(3) 517–529

© The Author(s) 2017

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164417717314

journals.sagepub.com/home/epm

Yong Luo¹

Abstract

Mplus is a powerful latent variable modeling software program that has become an increasingly popular choice for fitting complex item response theory models. In this short note, we demonstrate that the two-parameter logistic testlet model can be estimated as a constrained bifactor model in Mplus with three estimators encompassing limited- and full-information estimation methods.

Keywords

item response theory (IRT), testlet, limited-information estimation, full-information estimation, bifactor model

Introduction

The testlet model (e.g., Bradlow, Wang, & Wainer, 1999) is a popular approach used by item response theory (IRT) researchers and practitioners to address the issue of local item dependence (LID) caused by a cluster of items sharing the same stimulus, which can be a reading comprehension passage, a scenario, a graph, and so on. Estimation of this model has usually been conducted with full-information methods, including Markov chain Monte Carlo (MCMC) algorithm (e.g., Koziol, 2016; Li, Bolt, & Fu, 2006) and marginal maximum likelihood estimation (MMLE; e.g., Jiao, Wang, & He, 2013; Li, Li, & Wang, 2010), although Bolt (2005) showed that both full- and limited-information methods can be applied to multidimensional item response theory (MIRT; Reckase, 2009) models. As different estimation methods are

¹National Center for Assessment, Riyadh, Saudi Arabia

Corresponding Author:

Yong Luo, National Center for Assessment, King Khalid Road, Riyadh, 11537, Saudi Arabia.

Email: jackyluo1980@gmail.com

implemented in different software programs, researchers and practitioners may have to learn to use a software program with which they might not be familiar in order to use a particular estimation method. For example, SCORIGHT (Wang, Bradlow, & Wainer, 2004) and WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) are software programs implementing MCMC that are often used to estimate the testlet model; for implementation of MMLE, TESTFACT (Wilson, Wood, & Gibbons, 1991), ConQuest (Wu, Adams, Wilson, & Haldane, 2007), and the SAS NLMIXED procedure (SAS Institute, 2015) are popular choices. IRTPRO (Cai, Thissen, & du Toit, 2015) is a relatively recent software program that has the flexibility to implement both MCMC and MMLE estimation methods for the testlet model.

Mplus (L. Muthén & Muthén, 1998-2012) is a popular statistical software program for latent variable modeling that is known for its provision of a large number of estimators. In this article, we show that the two-parameter logistic (2PL) testlet model, a special case of the MIRT model, can be estimated in Mplus with different estimators, namely the robust weighted least square estimator (WLSMV; B. Muthén, du Toit, & Spisic, 1997), the maximum likelihood estimator with robust standard errors (MLR), and the Bayes estimator. Among these three estimators, the WLSMV estimator is a limited information estimation method (the dominant method used in structural equation modeling) due to its use of only bivariate information, and the other two are full-information estimation methods that are commonly used in IRT.

The 2PL Testlet Model

The 2PL testlet model is given as

$$p_j(\theta_i) = \frac{1}{1 + e^{-a_j(\theta_i - b_j - \gamma_{id(j)})}}, \quad (1)$$

where $p_j(\theta_i)$ is the probability of a correct response to item j for examinee i , θ_i is examinee i 's latent ability, a_j and b_j are the item discrimination and difficulty parameters, and $\gamma_{id(j)}$ is a person-specific testlet effect that is assumed to follow a distribution $N(0, \sigma_{\gamma_{id(j)}}^2)$.

The bifactor model (Gibbons & Hedeker, 1992) takes the following form:

$$p_j(\theta_i) = \frac{1}{1 + e^{-(a_{jg}\theta_g - b_j + a_{js}\theta_s)}}, \quad (2)$$

where a_{jg} and a_{js} are the item discrimination parameters on the general and specific factors for item j , b_j is the intercept parameter for item j , θ_{jg} and θ_{js} are examinee i 's latent abilities on the general and specific factors for item j .

It has been shown that the 2PL testlet model is a special case of the more general bifactor model in that the discrimination parameter on a testlet trait is constrained to be the product of $\sigma_{\gamma_{id(j)}}$ and the discrimination parameter on the primary trait (e.g., Li et al., 2006; Rijmen, 2010). Consequently, the 2PL testlet model can be estimated in Mplus as a bifactor model, with constraints imposed accordingly by setting the

loadings on the general factor and the testlet factor equal and freely estimating the testlet variances.

Estimation of the 2PL Testlet Model in Mplus

Estimating a 2PL testlet model in Mplus is analogous to conducting a CFA analysis with a restricted bifactor model. One constrains the factor loadings on the secondary factors representing testlet effects to be the same as the corresponding factor loadings on the general factor regardless of the estimator used. For model identification purposes, variance of the general factor is fixed to be 1 and variances of the secondary factors can be freely estimated.

To convert the estimated factor loadings and threshold parameters to the item parameters in Equation (1), different estimators require slightly different procedures. If WLSMV is the estimator (with the default Delta parameterization), the conversion can be directly conducted with the following two equations (McDonald, 1999):

$$a_j = \frac{1.702\lambda_j}{\sqrt{1 - \lambda_j' \phi \lambda_j}}, \quad (3)$$

$$b_j = \frac{-1.702\tau_j}{a_j \sqrt{1 - \lambda_j' \phi \lambda_j}}, \quad (4)$$

where a_j and b_j are the discrimination and difficulty parameters of item j , λ_j and τ_j are the vector of factor loading parameters and item threshold parameter, and ϕ is the factor covariance matrix. It should be noted that the constant 1.702 is used to convert item parameters from the normal metric to the logistic metric.

When the estimator is MLR or Bayes, in order to use Equations (3) and (4) we need to convert the factor loadings and thresholds estimated with MLR or Bayes to be on the same scale as those estimated with WLSMV. This conversion process requires the estimated unique variance of each item, which is not in the default output if either MLR or Bayes is used. In order to obtain the estimated item unique variance, we request Mplus to produce R^2 (the proportion of explained variance in each item), which can be subtracted from 1 to obtain the estimated item unique variance. According to Asparouhov and Muthén (2016), the following equations can be used to convert the factor loadings and item thresholds estimated using MLR or Bayes to those estimated using WLSMV with the default Delta parameterization:

$$\lambda_{jWLSMV} = \lambda_{jBayes/MLR} \sqrt{1 - R_j^2}, \quad (5)$$

$$\tau_{jWLSMV} = \tau_{jBayes/MLR} \sqrt{1 - R_j^2}, \quad (6)$$

where λ_{jWLSMV} and τ_{jWLSMV} are the factor loading and threshold for item j with WLSMV estimator, R_j^2 is the proportion of explained variance of item j by the latent

factor with either Bayes or MLR estimator, and $\lambda_{j\text{Bayes/MLR}}$ and $\tau_{j\text{Bayes/MLR}}$ are the factor loading and threshold for item j with either Bayes or MLR estimator.

A Demonstration With Simulated Data

In this section, we demonstrate with a simulated data set that the 2PL testlet model can be estimated in Mplus with three different estimators. We also show that the estimated item parameters are comparable to those with NOHARM (normal ogive harmonic analysis robust method; Fraser & McDonald, 1988), a well-known IRT program for estimating MIRT models (e.g., Knol & Berger, 1991; Svetina & Levy, 2016). To simulate data, we assumed that there was a 30-item reading comprehension test taken by 2,000 examinees randomly drawn from a standard normal distribution; these 30 items were evenly distributed across six testlets with each having five items; we further assumed that all six testlets had testlet variances equal to 1. Plugging item parameters listed in the columns under “TRUE” in Table 3 in Equation (1), we generated a data set for subsequent analyses conducted in this section.

We fit the 2PL testlet model to the generated data set using the three sets of Mplus syntaxes in Figures 1 to 3. Figure 1 shows the Mplus syntax for the WLSMV estimator. Under the **ANALYSIS** command (line 5), the estimator is specified to be WLSMV for didactic purpose, although WLSMV is the default estimator for CFA with categorical variables and this line can be omitted. We specify a general factor **f** (line 6) and six testlet factors **f1**, **f2**, **f3**, **f4**, **f5**, and **f6** (lines 7-12), and use the command **by** to indicate which indicators are linked to a specific factor. For example, **f by u1-u30** (line 6) means that the general factor **f** has 30 indicators named from **u1** to **u30**. We also use the asterisk (*) to free the default constraint in Mplus that forces the loading and threshold of the first indicator to be 1 and 0 (lines 6-12). By freeing this constraint and constraining the variance and mean of the general factor **f** to be 1 and 0 for model identification with the command **@** (line 13), loadings and thresholds of all items are freely estimated. We do not constrain the variance of the six testlet factors, which are usually freely estimated in testlet models. It should be noted that as the mean of any latent variable is constrained to be 0 at default in Mplus, the statement **[f@0]** is unnecessary but included here for illustration purpose. Consequently, although we do not use similar statements such as **[f1@0]** to explicitly constrain the means of the testlet factors to be 0, they are automatically set to 0 in Mplus.

Within the parenthesis (line 6) we name the loadings of the 30 items on the general factor **f** as **I1-I30**, which are referred to subsequently to impose the equality constraints inherent in the testlet model. For example, as the loadings of the first five items on **f1** (the first testlet factor) share the same name **I1-I5** (line 7), for each of the five items the factor loadings on the general factor and the first testlet factor are constrained to equality. Without such constraints, the model would be estimated as a general bifactor model other than the 2PL testlet model, which is a constrained bifactor model. To impose the orthogonal structures in the bifactor model, we constrain the correlation between the general factor **f** and the testlet factors **f1-f6** to be 0 with

1	TITLE: Testlet Model Estimation
2	DATA: FILE IS r1.txt;
3	VARIABLE: NAMES ARE u1-u30;
4	CATEGORICAL ARE u1-u30;
5	ANALYSIS: ESTIMATOR = WLSMV;
6	MODEL: f by u1-u30*(I1-I30);
7	f1 by u1-u5*(I1-I5);
8	f2 by u6-u10*(I6-I10);
9	f3 by u11-u15*(I11-I15);
10	f4 by u16-u20*(I16-I20);
11	f5 by u21-u25*(I21-I25);
12	f6 by u26-u30*(I26-I30);
13	f@1; [f@0];
14	f with f1@0 f2@0 f3@0 f4@0 f5@0 f6@0;
15	f1 with f2@0 f3@0 f4@0 f5@0 f6@0;
16	f2 with f3@0 f4@0 f5@0 f6@0;
17	f3 with f4@0 f5@0 f6@0;
18	f4 with f5@0 f6@0;
19	f5 with f6@0;

Figure 1. Mplus syntax for two-parameter logistic (2PL) testlet model estimation with robust weighted least square estimator (WLSMV).

the command **with** (line 14), and the correlation between testlet factors to be 0 (lines 15-19).

Figure 2 shows the Mplus syntax for the MLR estimator. As this set of syntax is largely the same as that provided in Figure 1, we focus on the differences. One noticeable difference between the Mplus syntax provided in Figure 1 and that in Figure 2 is that the standardized solution is requested in the OUTPUT section for the MLR estimator (line 21), which provides the R^2 estimate for each item. With WLSMV estimator Mplus automatically provides the R^2 estimate, so it is not necessary to request it in the OUTPUT section. In addition, in the Mplus syntax with MLR estimator, we specify the link function to be probit (line 6). As mentioned

1	TITLE: Testlet Model Estimation
2	DATA: FILE IS r1.txt;
3	VARIABLE: NAMES ARE u1-u30;
4	CATEGORICAL ARE u1-u30;
5	ANALYSIS: ESTIMATOR = MLR;
6	LINK=PROBIT;
7	MODEL: f by u1-u30*(I1-I30);
8	f1 by u1-u5*(I1-I5);
9	f2 by u6-u10*(I6-I10);
10	f3 by u11-u15*(I11-I15);
11	f4 by u16-u20*(I16-I20);
12	f5 by u21-u25*(I21-I25);
13	f6 by u26-u30*(I26-I30);
14	f@1; [f@0];
15	f with f1@0 f2@0 f3@0 f4@0 f5@0 f6@0;
16	f1 with f2@0 f3@0 f4@0 f5@0 f6@0;
17	f2 with f3@0 f4@0 f5@0 f6@0;
18	f3 with f4@0 f5@0 f6@0;
19	f4 with f5@0 f6@0;
20	f5 with f6@0;
21	OUTPUT: STANDARDIZED;

Figure 2. Mplus syntax for two-parameter logistic (2PL) testlet model estimation with maximum likelihood estimator with robust standard errors (MLR).

earlier, the default link function with MLR estimator in Mplus is logit, and we change it to probit to be consistent with WLSMV and Bayes.

Figure 3 shows the Mplus syntax for the Bayes estimator. Same as the syntax for the MLR estimator, the standardized solution is also requested in the OUTPUT section for the Bayes estimator, as well as **TECH8** that provided various diagnostic information regarding model convergence (line 23). The Bayes estimator in Mplus uses an MCMC algorithm based on the Gibbs sampler, and in the current syntax we specify four processors to be used (line 6) to run four parallel chains (line 8). In

1	TITLE: Testlet Model Estimation
2	DATA: FILE IS r1.txt;
3	VARIABLE: NAMES ARE u1-u30;
4	CATEGORICAL ARE u1-u30;
5	ANALYSIS: ESTIMATOR = BAYES;
6	PROCESSORS = 4;
7	BITERATIONS = (20000);
8	CHAINS = 4;
9	MODEL: f by u1-u30* (I1-I30);
10	f1 by u1-u5* (I1-I5);
11	f2 by u6-u10* (I6-I10);
12	f3 by u11-u15* (I11-I15);
13	f4 by u16-u20* (I16-I20);
14	f5 by u21-u25* (I21-I25);
15	f6 by u26-u30* (I26-I30);
16	f@1; [f@0];
17	f with f1@0 f2@0 f3@0 f4@0 f5@0 f6@0;
18	f1 with f2@0 f3@0 f4@0 f5@0 f6@0;
19	f2 with f3@0 f4@0 f5@0 f6@0;
20	f3 with f4@0 f5@0 f6@0;
21	f4 with f5@0 f6@0;
22	f5 with f6@0;
23	OUTPUT: STANDARDIZED TECH8;

Figure 3. Mplus syntax for two-parameter logistic (2PL) testlet model estimation with Bayes estimator.

addition, we use the command **BITERATIONS** to specify the minimum iterations for each chain to be 20,000 (line 7). Mplus uses the Gelman and Rubin’s convergence diagnostic (Gelman & Rubin, 1992) to judge model convergence, and in the current case if model convergence is still not reached after 20,000 iterations, Mplus will continue running the four MCMC chains until the model has converged.

Table 1. Testlet Variances Estimates.

Estimator	T1	T2	T3	T4	T5	T6
WLSMV	1.246	1.222	1.079	0.994	1.01	1.233
MLR	1.282	1.252	1.089	1.015	1	1.229
Bayes	1.282	1.238	1.092	1.005	0.991	1.233
NOHARM	1.331	1.245	1.081	0.965	1.025	1.220

Note. NOHARM = normal ogive harmonic analysis robust method; WLSMV = robust weighted least square estimator; MLR = maximum likelihood estimator with robust standard errors.

Table 1 lists the testlet factor variance estimates in Mplus with the three estimators and in NOHARM. As the variance of the general factor was constrained to follow a standard normal distribution for model identification, a practice also implemented in IRT framework, these estimates can be directly interpreted as testlet variance estimates without additional transformation. As can be seen, all three estimators in Mplus produce factor variance estimates close to the generating value of one and similar to those in NOHARM.

Table 2 lists the factor loadings, item thresholds, and R^2 estimates with the three estimators from the Mplus output. As can be seen, the factor loading and item threshold estimates with WLSMV estimator are not on the same scale as those estimated with MLR and Bayes estimators.

To convert the factor loadings and item thresholds to item discrimination and difficulty parameters, we directly apply Equations (3) and (4) to the estimates with WLSMV estimator and, for the estimates with MLR and Bayes estimator, we first apply Equations (5) and (6) to convert the estimates on the same scale as with WLSMV estimator, and plug in these results in Equations (2) and (3) to obtain the corresponding parameters in the IRT framework. In the following, we use the first item as an example to illustrate how to convert the factor loadings and item thresholds with MLR estimator to their IRT analogs. For this item,

$$\lambda_{jWLSMV} = \lambda_{MLR} \sqrt{1 - R_j^2} = 0.63 * \sqrt{1 - 0.485} = 0.461$$

$$\tau_{jWLSMV} = \tau_{MLR} \sqrt{1 - R_j^2} = -1.032 * \sqrt{1 - 0.485} = -0.741$$

As this item belongs to the first testlet and its loadings on the other testlet factors are 0, its factor loading vector λ_1 is

$$\begin{bmatrix} 0.461 \\ 0.461 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

Table 2. Factor Loading and Item Threshold Estimates.

Item	WLSMV			MLR			Bayes		
	λ	τ	R^2	λ	τ	R^2	λ	τ	R^2
1	0.462	−0.74	0.48	0.643	−1.032	0.485	0.646	−1.034	0.489
2	0.284	−0.381	0.181	0.316	−0.422	0.186	0.318	−0.423	0.188
3	0.368	0.5	0.304	0.43	0.596	0.297	0.433	0.597	0.3
4	0.417	0.883	0.391	0.491	1.1	0.355	0.492	1.1	0.356
5	0.419	−0.432	0.395	0.557	−0.563	0.415	0.559	−0.563	0.416
6	0.383	0.015	0.326	0.462	0.019	0.325	0.465	0.019	0.326
7	0.427	−0.341	0.405	0.547	−0.441	0.402	0.55	−0.442	0.404
8	0.43	−0.261	0.412	0.553	−0.339	0.407	0.557	−0.34	0.411
9	0.261	0.499	0.152	0.292	0.545	0.161	0.293	0.546	0.162
10	0.388	0.04	0.335	0.456	0.05	0.319	0.459	0.05	0.321
11	0.357	0.09	0.265	0.421	0.105	0.271	0.423	0.106	0.273
12	0.413	−0.088	0.354	0.495	−0.108	0.339	0.495	−0.108	0.34
13	0.518	0.958	0.557	0.778	1.44	0.558	0.775	1.438	0.558
14	0.451	−0.247	0.423	0.587	−0.323	0.419	0.59	−0.325	0.423
15	0.434	0.138	0.392	0.572	0.18	0.406	0.572	0.18	0.407
16	0.389	0.14	0.302	0.449	0.167	0.289	0.452	0.167	0.291
17	0.449	−0.565	0.402	0.569	−0.724	0.395	0.573	−0.724	0.397
18	0.368	0.402	0.27	0.457	0.479	0.296	0.46	0.478	0.298
19	0.41	0.364	0.335	0.486	0.442	0.323	0.49	0.443	0.326
20	0.263	−0.011	0.138	0.292	−0.012	0.147	0.294	−0.013	0.148
21	0.333	−0.141	0.223	0.377	−0.16	0.221	0.378	−0.16	0.223
22	0.428	0.469	0.369	0.565	0.603	0.389	0.568	0.603	0.392
23	0.438	0.055	0.386	0.55	0.072	0.377	0.555	0.072	0.381
24	0.39	−0.466	0.305	0.45	−0.553	0.289	0.452	−0.553	0.29
25	0.311	0.26	0.194	0.351	0.291	0.197	0.352	0.291	0.199
26	0.407	0.242	0.37	0.512	0.304	0.369	0.514	0.305	0.371
27	0.343	−0.025	0.262	0.382	−0.028	0.245	0.383	−0.028	0.247
28	0.318	−0.294	0.225	0.366	−0.335	0.23	0.366	−0.335	0.231
29	0.396	−0.048	0.35	0.49	−0.059	0.349	0.492	−0.058	0.351
30	0.355	−0.176	0.282	0.437	−0.21	0.299	0.438	−0.21	0.3

Note. NOHARM = normal ogive harmonic analysis robust method; WLSMV = robust weighted least square estimator; MLR = maximum likelihood estimator with robust standard errors.

and the factor covariance matrix Φ is the following diagonal matrix with the estimated factor variance (extracted from Table 1) on the diagonal:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.282 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.252 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.089 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.015 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.229 \end{bmatrix}.$$

Table 3. True and Converted Item Parameter Estimates.

Item	True		NOHARM		WLSMV		MLR		Bayes	
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>A</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
1	1.173	−1.553	1.117	−1.598	1.090	−1.602	1.095	−1.605	1.097	−1.601
2	0.612	−1.288	0.529	−1.357	0.534	−1.342	0.538	−1.335	0.541	−1.330
3	0.674	1.441	0.735	1.387	0.751	1.359	0.732	1.386	0.737	1.379
4	1.156	1.859	0.936	2.098	0.909	2.118	0.836	2.240	0.837	2.236
5	1.058	−0.904	0.912	−1.041	0.916	−1.031	0.947	−1.011	0.952	−1.007
6	0.693	0.054	0.791	0.039	0.794	0.039	0.786	0.041	0.792	0.041
7	0.805	−0.881	0.934	−0.805	0.942	−0.799	0.932	−0.806	0.936	−0.804
8	0.95	−0.617	0.943	−0.612	0.953	−0.607	0.942	−0.613	0.946	−0.610
9	0.512	1.887	0.473	1.942	0.482	1.912	0.497	1.866	0.498	1.863
10	0.883	0.094	0.812	0.103	0.810	0.103	0.776	0.110	0.781	0.109
11	0.781	0.197	0.710	0.252	0.709	0.252	0.716	0.249	0.720	0.251
12	0.962	−0.189	0.897	−0.209	0.875	−0.213	0.842	−0.218	0.841	−0.218
13	1.212	1.888	1.312	1.859	1.326	1.849	1.325	1.851	1.315	1.855
14	0.897	−0.5	1.004	−0.549	1.010	−0.548	0.998	−0.550	1.002	−0.551
15	0.94	0.271	0.938	0.321	0.947	0.318	0.974	0.315	0.973	0.315
16	0.761	0.345	0.807	0.354	0.792	0.360	0.764	0.372	0.769	0.369
17	0.977	−1.245	0.982	−1.260	0.988	−1.258	0.968	−1.272	0.975	−1.264
18	0.697	1.299	0.728	1.093	0.733	1.092	0.778	1.048	0.783	1.039
19	0.899	0.83	0.861	0.881	0.856	0.888	0.827	0.909	0.833	0.904
20	0.581	0.061	0.485	−0.042	0.482	−0.042	0.497	−0.041	0.500	−0.044
21	0.663	−0.406	0.633	−0.427	0.643	−0.423	0.642	−0.424	0.643	−0.423
22	0.837	1.09	0.916	1.099	0.916	1.096	0.962	1.067	0.966	1.062
23	0.81	0.014	0.950	0.125	0.951	0.126	0.936	0.131	0.944	0.130
24	0.775	−1.062	0.788	−1.207	0.797	−1.195	0.765	−1.229	0.769	−1.223
25	0.495	0.891	0.584	0.843	0.590	0.836	0.598	0.829	0.599	0.827
26	0.97	0.625	0.877	0.592	0.873	0.595	0.871	0.594	0.875	0.593
27	0.615	−0.171	0.684	−0.072	0.680	−0.073	0.650	−0.073	0.652	−0.073
28	0.7	−0.808	0.616	−0.923	0.615	−0.925	0.623	−0.915	0.623	−0.915
29	0.816	−0.119	0.834	−0.120	0.836	−0.121	0.834	−0.120	0.837	−0.118
30	0.774	−0.426	0.710	−0.499	0.713	−0.496	0.743	−0.481	0.745	−0.479

Note. NOHARM = normal ogive harmonic analysis robust method; WLSMV = robust weighted least square estimator; MLR = maximum likelihood estimator with robust standard errors.

Plugging the above values into Equations (3) and (4), we obtain the item parameter values for the first item listed under “MLR” in Table 3. As can be seen, the item parameter estimates with the three estimators in Mplus are similar to those in NOHARM and close to the true generating item parameter values.

Conclusions

The testlet model is a popular statistical model frequently used in the psychometric literature to address LID due to a cluster of items sharing the same stimulus, and it

has been estimated with either MCMC algorithm or MMLE implemented in various software programs. To our knowledge there are no tutorials in the psychometric literature that show how to estimate the testlet models in Mplus, a popular latent variable modeling software program that has been used to estimate various complex IRT models (e.g., Finch & Bolin, 2017; Huggins-Manley & Algina, 2015). As a result, many Mplus users may have to resort to other statistical software programs with which they are not familiar if they would like to investigate possible LID in their data with the testlet model. This article adds to the current literature by showing that Mplus can be a viable tool for the estimation of testlet models with its provision of both limited- and full-information-based estimation methods.

Specifically, we demonstrated with a simulated data set that three different estimators used in Mplus encompassing both limited- and full-information estimation methods, namely WLSMV, MLR, and Bayes, can be used to estimate the 2PL testlet model (while Mplus can estimate the three-parameter logistic unidimensional IRT model, its testlet analog cannot be estimated), and the results were comparable to NOHARM, a well-established MIRT software program. We did not focus on the one-parameter logistic (1PL) testlet model, but as a special case of the 2PL testlet model it can be estimated with slight modification of the Mplus syntaxes provided in Figures 1 to 3. In the appendix, we provide example Mplus syntax for the 1PL testlet model with WLSMV estimator. As the 1PL testlet model differs from the 2PL testlet model in that all the item discrimination parameters are the same, we constrain the loadings of all items on both the general factor and the specific testlet factor to equality by naming them as **I1** (lines 6-12).

As the current article is of didactic nature, only one data set was simulated and consequently, we did not compute bias, standard error, and root mean square error, indices commonly used in simulation studies with multiple replications to evaluate parameter recovery. For applied researchers and practitioners who are Mplus users interested in applying the testlet model, the provided Mplus codes are readily usable and there is no need to migrate to other IRT statistical software programs that may take time and resources to learn. Methodological researchers may also benefit from the provided Mplus syntaxes, which can be used for simulation studies to investigate statistical and psychometric properties of the testlet model, and as the three estimators in Mplus illustrated in this article encompass different estimation methods, comparison studies of estimation methods can be easily conducted within Mplus without using additional software programs.

Appendix

Mplus Syntax for IPL Testlet Model With WLSMV Estimator

```

1  TITLE: IPL Testlet Model Estimation
2  DATA: FILE IS r1.txt;
3  VARIABLE: NAMES ARE u1-u30;
4  CATEGORICAL ARE u1-u30;
5  ANALYSIS: ESTIMATOR = WLSMV;
6  MODEL: f by u1-u30* (11);
7  f1 by u1-u5* (11);
8  f2 by u6-u10* (11);
9  f3 by u11-u15* (11);
10 f4 by u16-u20* (11);
11 f5 by u21-u25* (11);
12 f6 by u26-u30* (11);
13 f@1; [f@0];
14 f with f1@0 f2@0 f3@0 f4@0 f5@0 f6@0;
15 f1 with f2@0 f3@0 f4@0 f5@0 f6@0;
16 f2 with f3@0 f4@0 f5@0 f6@0;
17 f3 with f4@0 f5@0 f6@0;
18 f4 with f5@0 f6@0;
19 f5 with f6@0;

```

Note. WLSMV = robust weighted least square estimator; IPL = one-parameter logistic.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Asparouhov, T., & Muthén, B. (2016). *IRT in Mplus* (Technical report). Los Angeles, CA: Muthén & Muthén.
- Bolt, D. M. (2005). Limited and full-information IRT estimation. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics* (pp. 27-71). Mahwah, NJ: Lawrence Erlbaum.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64, 153-168.
- Cai, L., Thissen, D., & du Toit, S. H. C. (2015). *IRTPRO for Windows* [Computer software]. Lincolnwood, IL: Scientific Software International.
- Finch, H., & Bolin, J. (2017). *Multilevel modeling using Mplus*. Boca Raton, FL: CRC.

- Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research*, 23, 267-269.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.
- Gibbons, R. D., & Hedeker D. (1992). Full-information item bifactor analysis. *Psychometrika*, 57, 423-436.
- Huggins-Manley, A. C., & Algina, J. (2015). The partial credit model and generalized partial credit model as constrained nominal response models, with applications in Mplus. *Structural Equation Modeling*, 22, 308-318.
- Jiao, H., Wang, S., & He, W. (2013). Estimation methods for one-parameter testlet models. *Journal of Educational Measurement*, 50, 186-203.
- Knol, D. L., & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, 26, 457-477.
- Kozioł, N. A. (2016). Parameter recovery and classification accuracy under conditions of testlet dependency: A comparison of the traditional 2PL, testlet, and bi-factor models. *Applied Measurement in Education*, 29, 184-195.
- Li, Y., Bolt, D. M., & Fu, J. (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement*, 30, 3-21.
- Li, Y., Li, S., & Wang, L. (2010). *Application of a general polytomous testlet model to the reading section of a large-scale English language assessment* (ETS RR-10-21). Princeton, NJ: Educational Testing Service.
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.
- McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Lawrence Erlbaum.
- Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript.
- Muthén, L., & Muthén, B. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Reckase, M. (2009). *Multidimensional item response theory* (Vol. 150). New York, NY: Springer.
- Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47, 361-372.
- SAS Institute. (2015). *SAS/STAT user's guide* (Version 9.4). Cary, NC: Author.
- Svetina, D., & Levy, R. (2016). Dimensionality in compensatory MIRT when complex structure exists: Evaluation of DETECT and NOHARM. *Journal of Experimental Education*, 84, 398-420.
- Wang, X., Bradlow, E. T., & Wainer, H. (2004). *User's guide for SCORIGHT (Version 3.0): A computer program for scoring tests built of testlets including a module for covariate analysis* (ETS Research Report RR 04-49). Princeton, NJ: Educational Testing Service.
- Wilson, D. T., Wood, R., & Gibbons, R. D. (1991). *TESTFACT: Test scoring, item statistics, and item factor analysis*. Skokie, IL: Scientific Software International.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). *ACER ConQuest 2.0: General item response modelling software* [computer program manual]. Camberwell, Victoria: Australian Council for Educational Research Press.