

**The Vuong-Lo-Mendell-Rubin Test for Latent Class and Latent Profile  
Analysis: A Note on the Different Implementations in Mplus and  
LatentGOLD**

Jeroen K. Vermunt

Tilburg University, Tilburg, The Netherlands

*Corresponding author's contact address:* Jeroen K. Vermunt, Tilburg School of Social and Behavioral Sciences, Department of Methodology and Statistics, PO box 90153, 5000 LE Tilburg, The Netherlands. E-mail: j.k.vermunt@uvt.nl.

*Published in 2024 in Methodology, European Journal of Research Methods for the Behavioral and Social Sciences*

# **The Vuong-Lo-Mendell-Rubin Test for Latent Class and Latent Profile Analysis: A Note on the Different Implementations in Mplus and LatentGOLD**

## **Abstract**

Mplus and LatentGOLD implement the Vuong-Lo-Mendell-Rubin test (comparing models with  $K$  and  $K+1$  latent classes) in slightly different manners. While LatentGOLD uses the formulae from Vuong (1989), Mplus replaces the standard parameter variance-covariance matrix by its robust version. Our small simulation study showed why such a seemingly small difference may sometimes yield rather different results. The main finding is that the Mplus approximation of the distribution of the likelihood-ratio statistic is much more data dependent than the LatentGOLD one. This data dependency is stronger when the true model serves as the null hypothesis ( $H_0$ ) with  $K$  classes than when it serves as the alternative hypothesis ( $H_1$ ) with  $K+1$  classes, and it is also stronger for low class separation than for high class separation. Another important finding is that neither of the two implementations yield uniformly distributed p-values under the correct null hypothesis, indicating this test is not the best model selection tool in mixture modeling.

*Keywords: Class enumeration, mixture modeling, likelihood-ratio test, nested models, VLMR test.*

Since version 2.12, the Mplus program contains an option to output the Vuong-Lo-Mendell-Rubin (VLMR) test for the comparison of mixture models with  $K$  and  $K+1$  classes (Muthén and Muthén, 2002). This test is based on the work by Vuong (1989) who proposed a generalized likelihood-ratio (LR) test for comparing two models in situations in which the standard LR test is not valid. Lo, Mendell, and Rubin (2001) proposed applying Vuong's LR test in the context of mixture models. More specifically, they showed how it can be used for comparing a  $K$ -class and a  $K+1$ -class mixture model of univariate normal distributions. Because Mplus implements the VLMR test for any type of mixture model, it is commonly used by Mplus users as an alternative to the computationally more demanding bootstrap likelihood-ratio test (BLRT) in the context of latent class analysis (LCA), latent profile analysis (LPA), and mixture growth modeling. A non-significant result indicates the model with  $K+1$  classes does not fit better than the model with  $K$  classes, implying the  $K$ -class model can be retained.

The tutorial on the UCLA webpage "<https://stats.idre.ucla.edu/mplus/dac/latent-class-analysis/>" illustrates both the VLMR test and the BLRT provided by Mplus using a LCA with 9 dichotomous indicators. When testing the 2-class model against the 3-class model, the authors obtained a LR value of 39.025, for which the VLMR test and the BLRT yielded p-values of .15 and .00, respectively. These results did not only contradict one another, the non-significant p-value of the VLMR test for the rather large LR value of 39.025 with only 10 parameters difference was also somewhat counterintuitive. The tutorial authors also expressed some doubts about the VLMR test result, and therefore proposed using the 3-class model as the final model (thus following the BLRT result).

On popular request, the VLMR test was implemented in LatentGOLD version 6.0 (Vermunt and Magidson, 2021). However, when comparing LatentGOLD's results with those

reported by Mplus for the data set on the UCLA website, we noticed that our own calculations yielded a highly significant p-value ( $p=.00$ ), differing substantially from the Mplus result ( $p=.15$ ). Mplus also reports the mean and standard deviation of the estimated VLMR distribution, which for the application concerned yielded a mean of 20.26 and a standard deviation of 22.22, while LatentGOLD reports values of 11.80 and 7.49, respectively. This shows that the two programs are using rather different distributions to obtain the p-value corresponding to the observed VLMR value. Fortunately, we were able to exactly reproduce the Mplus results with an alternative implementation of the Vuong test; that is, by replacing the negative inverse Hessian (the non-robust estimator of the variance-covariance matrix of the model parameters) by its robust or sandwich estimator. Note that Mplus requires using the MLR (maximum likelihood robust) estimator when requesting the VLMR test with the TECH11 option, which hinted us in this direction. Though the Mplus developers may have had good reasons for using this modification of the Vuong test, we have not been able to find a theoretical justification for this choice in the literature.

Let us look in more detail into the Vuong test of interest, which he referred to as the LR test for nested or overlapping models (note that he proposed another test for non-nested or non-overlapping models). According to Vuong (1989), in such situations (under some regularity conditions) the asymptotic distribution of the LR statistic is a weighted sum of  $\chi_1^2$  random variables, where the (possibly negative) weights are the eigenvalues of a matrix  $\mathbf{W}_{Vuong}$ . This matrix is defined as follows:

$$\mathbf{W}_{Vuong} = \begin{bmatrix} -\mathbf{B}_{H1}\mathbf{A}_{H1}^{-1} & -\mathbf{B}_{H1H0}\mathbf{A}_{H1}^{-1} \\ \mathbf{B}_{H1H0}'\mathbf{A}_{H0}^{-1} & \mathbf{B}_{H0}\mathbf{A}_{H0}^{-1} \end{bmatrix},$$

where  $\mathbf{A}_{H1}$  and  $\mathbf{A}_{H0}$  are matrices of second derivatives of the log-likelihood of the model serving as alternative hypothesis (H1) and as null hypothesis (H0) (in mixture modeling, the K+1 and K-

class model), respectively, and  $\mathbf{B}_{H1}$ ,  $\mathbf{B}_{H0}$ , and  $\mathbf{B}_{H1H0}$  are matrices containing sums across observations of the cross-products of the first derivatives of the individual log-likelihood contributions of the H1 model ( $\mathbf{B}_{H1}$ ), of the H0 model ( $\mathbf{B}_{H0}$ ), and of the H0 by H1 model ( $\mathbf{B}_{H1H0}$ ). This is the formulation used by LatentGOLD 6.0 (Vermunt and Magidson, 2021). The Mplus implementation in the following:

$$\mathbf{W}_{Mplus} = \begin{bmatrix} \mathbf{B}_{H1} \mathbf{V}_{H1}^{-1} & \mathbf{B}_{H1H0} \mathbf{V}_{H1}^{-1} \\ -\mathbf{B}_{H1H0}' \mathbf{V}_{H0}^{-1} & -\mathbf{B}_{H0} \mathbf{V}_{H0}^{-1} \end{bmatrix},$$

where  $\mathbf{V}_{H1}^{-1} = \mathbf{A}_{H1}^{-1} \mathbf{B}_{H1} \mathbf{A}_{H1}^{-1}$  and  $\mathbf{V}_{H0}^{-1} = \mathbf{A}_{H0}^{-1} \mathbf{B}_{H0} \mathbf{A}_{H0}^{-1}$ . That is, Mplus replaces minus the inverse Hessian  $-\mathbf{A}_{H1}^{-1}$  and  $-\mathbf{A}_{H0}^{-1}$  by the robust variance estimators  $\mathbf{V}_{H1}^{-1}$  and  $\mathbf{V}_{H0}^{-1}$ .

The sum of the eigenvalues of  $\mathbf{W}_{Vuong}$  (or of  $\mathbf{W}_{Mplus}$  in Mplus) yields the mean of the (estimated) distribution of the VLMR statistic, whereas the square root of twice the sum of the squared eigenvalues yields its standard deviation. The p-value for the observed LR value is obtained using the method proposed by Imhof (1961).

Note that applications of the general Vuong test described above, as well as of variants for non-nested and non-overlapping models, have also been proposed in structural equation modeling (Merkle, You, and Preacher, 2016) and item response theory modeling (Schneider et al., 2020). Since the Vuong tests implemented in LatentGOLD are also applicable to structural equation and item response theory models, we were able to confirm that the LatentGOLD results match those obtained with the R package nonnest2 (Merkle and You, 2018). It should be noted that the Vuong test may also be applied to optimization functions beyond maximum likelihood (Golden, 2003).

A small simulation study was performed with the aim to explore the consequences of the different implementation of the VLMR test in Mplus and LatentGOLD. Our simulation was not

meant to show that one method is better than the other, but to gain some understanding on why the two methods may give different results; that is, on why they may give different p-values.

## Method

For our small simulation study, we used the LCA and LPA conditions of the well-known simulation study by Nylund, Asparouhov and Muthén (2007) as a starting point. More specifically, from their Table 2, we took the LCA and LPA populations with 8 items, 4 classes, and equal class sizes. The entropy R-squared values were .79 and .88 for the LCA and LPA population, respectively, showing classes are well separated in both conditions. We also consider using their LCA and LPA populations with 15 items, but these had very large entropy R-squared values (.98 and 1.00, respectively), which we thought would be less interesting settings for a mixture model simulation. Because we also wanted to compare the two versions of the VLMR test in a condition with less well separated classes, we took the 3-class maximum likelihood solution obtained with the data set from the UCLA website as our third condition. This 9-item population model has unequal class sizes and an entropy R-squared value of .44. The sample size was set to 1000 in all three conditions, and we run 1000 replications per condition.

While Nylund et al. (2007) focused on the type I error rate and the power for the VLMR test for a given alpha level, we investigated:

- 1) The sampling distribution of the mean and the standard deviation of the estimated distribution of the VLMR statistic. As explained above, these are simple functions of the eigenvalues of  $\mathbf{W}_{Vuong}$  (or  $\mathbf{W}_{Mplus}$ ). Ideally these should not vary too much across replication samples.
- 2) The full sampling distribution of the p-values. Ideally, this distribution should be close to uniform when testing the true model (i.e., when  $H_0$  is true).

The simulation study was performed using the Syntax version of LatentGOLD 6.0, which also allows obtaining the Mplus version of the VLMR test by requesting robust standard errors and adding the keyword “mplus” to the list of output options. When running a LCA or LPA for a range of classes at once, VLMR statistics are obtained automatically as part of the output. The appendix shows the syntax used to generate a data set, as well as the syntax used to run the models for a simulated data set.

## Results

Table 1 presents the results we obtained when testing the true model with K classes (as H0) against the alternative model with K+1 classes (as H1). This table provides information on the sampling distribution of the estimated mean and estimated standard deviation of the VLMR distribution used to obtain the p-values (“Mean of VLMR distribution” and “StdDev of VLMR distribution”), as well as on the sampling distribution of the p-values themselves. For these quantities, we report a series of percentiles, the mean, and the standard deviation across 1000 simulation replications.

As can be seen, “Mean of VLMR distribution” and “StdDev of VLMR distribution” vary considerably across replications (see, for example, the difference between the 5<sup>th</sup> and 95<sup>th</sup> percentile and the value reported in the “StdDev” column). This variation is largest for the LCA with low class separation (“LCA-3”) followed by the LPA (“LPA-4”) and the LCA with high separation (“LCA-4”). But more importantly, the variation is much larger for Mplus than for LatentGOLD. For example, for the “LCA-3” condition, the 5<sup>th</sup> and 95<sup>th</sup> percentile of “Mean of VLMR distribution” equal 2.83 and 14.81 for LatentGOLD, whereas these equal -14.34 and 57.14 for Mplus. A similar pattern can be observed for “StdDev of VLMR distribution”. What can also be seen is that the mean of these quantities across replications (i.e., in the “Mean”

column) is much larger for Mplus than for LatentGOLD, which shows that (on average) the two programs use rather different distributions for obtaining the VLMR p-values. Moreover, the fact that percentiles and “VLMR p-values” do not match with one another shows the p-values are clearly not uniformly distributed in the three investigated conditions. This applies both to Mplus and LatentGOLD, though the Mplus p-values are closer to uniform than those of LatentGOLD.

Table 2 presents the same measures as Table 1, but now for the VLMR test of the model with K-1 classes (as H0) against the true model with K classes (as H1). As can be seen, compared to what we saw in Table 1, the sampling variation of “Mean of VLMR distribution” and “StdDev of VLMR distribution” is rather small with LatentGOLD, though still somewhat larger in the low separation condition (“LCA-3”). Again, Mplus shows larger sampling variation than LatentGOLD, and this difference is largest in the low separation condition. The means of “Mean of VLMR distribution” and “StdDev of VLMR distribution” are again (much) larger for Mplus than for LatentGOLD, which shows that also for this test the two VLMR versions use rather different distributions for obtaining the p-values. The p-value is always 0 in the conditions with a high-class separation, which corresponds to a power of 1.0 as was also reported by Nylund et al. (2007). In the low separation condition, we see that the p-value is smaller than .05 up to 95<sup>th</sup> percentile for LatentGOLD but already larger than .05 from the 75<sup>th</sup> percentile for Mplus.



1

Table 1: Results of the simulation studies (1000 replications): Characteristics of the estimated VLMR distribution across replications when comparing the true K-class model (H0) with the K+1-class model (H1)													
Model	Method	Measure	Percentiles									Mean	StdDev
			1%	5%	10%	25%	50%	75%	90%	95%	99%		
LCA-4	LatentGOLD	Mean VLMR distribution	3.11	4.33	5.10	6.13	7.32	8.91	10.69	11.66	16.21	7.65	2.45
Nylund et. al		StdDev VLMR distribution	2.86	3.50	3.77	4.17	4.81	5.91	7.40	8.94	14.19	5.35	2.11
		p-value of VLMR	0.00	0.00	0.00	0.02	0.04	0.11	0.19	0.27	0.43	0.08	0.09
	Mplus	Mean VLMR distribution	3.08	5.00	5.91	7.41	10.32	16.21	25.85	39.48	99.64	15.38	20.49
		StdDev VLMR distribution	3.50	4.41	4.86	6.19	8.67	14.95	27.48	43.46	129.88	15.63	27.74
		p-value of VLMR	0.00	0.01	0.02	0.07	0.17	0.34	0.54	0.63	0.80	0.23	0.20
LPA-4	LatentGOLD	Mean VLMR distribution	0.97	5.54	7.02	8.76	10.21	12.14	14.45	16.63	21.84	10.62	3.83
Nylund et. al		StdDev VLMR distribution	1.50	4.22	4.72	5.46	6.45	8.49	11.23	14.12	20.94	7.46	3.86
		p-value of VLMR	0.00	0.01	0.01	0.03	0.08	0.16	0.26	0.34	0.49	0.12	0.11
	Mplus	Mean VLMR distribution	0.90	9.09	11.45	15.12	21.36	37.12	67.01	105.25	265.84	36.85	67.50
		StdDev VLMR distribution	1.92	7.36	8.74	12.35	19.83	37.80	79.49	132.81	358.43	41.31	94.10
		p-value of VLMR	0.00	0.06	0.11	0.23	0.39	0.59	0.73	0.81	0.90	0.41	0.23
LCA-3	LatentGOLD	Mean VLMR distribution	-1.01	2.83	4.02	6.00	7.84	9.74	12.35	14.81	19.00	8.08	4.52
UCLA website		StdDev VLMR distribution	3.44	4.27	4.68	5.43	6.53	8.09	10.52	13.36	20.30	7.46	4.68
		p-value of VLMR	0.00	0.00	0.01	0.02	0.05	0.12	0.22	0.29	0.46	0.09	0.10
	Mplus	Mean VLMR distribution	-99.46	-14.34	-1.44	6.42	11.80	20.37	37.81	57.14	168.96	22.74	144.44
		StdDev VLMR distribution	5.34	7.39	8.56	11.67	17.83	30.92	59.91	96.76	280.64	42.70	201.37
		p-value of VLMR	0.00	0.02	0.04	0.12	0.24	0.43	0.63	0.72	0.88	0.29	0.22

2

3

Table 2: Results of the simulation studies (1000 replications): Characteristics of the estimated VLMR distribution across replications when comparing the K-1-class model (H0) with the true K-class model (H1)

Model	Method	Measure	Percentiles									Mean	StdDev
			1%	5%	10%	25%	50%	75%	90%	95%	99%		
LCA-4	LatentGOLD	Mean VLMR distribution	8.28	8.48	8.62	8.74	8.89	9.02	9.26	9.87	10.16	8.93	0.36
Nylund et. al		StdDev VLMR distribution	5.24	5.31	5.36	5.45	5.56	5.67	5.78	5.86	6.09	5.57	0.17
		p-value of VLMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mplus	Mean VLMR distribution	2.89	4.13	4.69	5.40	6.05	6.64	7.26	7.68	8.51	6.00	1.13
		StdDev VLMR distribution	8.42	8.73	8.92	9.35	9.78	10.32	11.03	11.66	13.51	9.93	0.97
		p-value of VLMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LPA-4	LatentGOLD	Mean VLMR distribution	8.90	9.25	9.39	9.64	9.90	10.12	10.33	10.44	10.71	9.87	0.37
Nylund et. al		StdDev VLMR distribution	6.05	6.18	6.24	6.38	6.53	6.69	6.85	6.97	7.30	6.54	0.25
		p-value of VLMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mplus	Mean VLMR distribution	-2.90	-0.45	0.60	1.91	3.11	4.26	5.05	5.57	6.31	2.93	1.85
		StdDev VLMR distribution	15.49	16.06	16.76	17.69	18.95	20.37	21.84	22.97	26.62	19.20	2.18
		p-value of VLMR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LCA-3	LatentGOLD	Mean VLMR distribution	1.29	5.72	6.75	7.99	9.35	10.74	12.15	13.81	18.85	9.47	2.99
UCLA website		StdDev VLMR distribution	4.54	4.99	5.42	6.18	7.06	8.23	9.73	11.94	20.63	7.64	2.98
		p-value of VLMR	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.15	0.01	0.02
	Mplus	Mean VLMR distribution	-91.62	-20.37	-8.66	3.14	9.04	14.40	25.32	36.73	123.94	8.76	38.40
		StdDev VLMR distribution	6.19	7.77	9.65	13.75	19.97	31.53	52.61	86.14	283.08	31.65	50.41
		p-value of VLMR	0.00	0.00	0.00	0.00	0.01	0.06	0.22	0.36	0.65	0.07	0.13

4

## Conclusion and Discussion

In this paper, we explained how the Mplus and LatentGOLD implementations of the VLMR test differ from one another. While LatentGOLD uses the formulae from Vuong (1989) and Lo, Mendell, and Rubin (2001), Mplus uses slightly modified formulae where the standard non-robust variance-covariance matrix of the parameters is replaced by its robust version.

We performed a small simulation study to explore the consequences of this seemingly minor difference. Our simulation was not meant to show that one method is better than the other, but to gain some understanding on why the two methods may give different results and to raise awareness regarding these differences among potential users.

In the simulation study we saw much larger variation of characteristics of the estimated sampling distribution (its mean and its standard error) across simulated data set with Mplus than with LatentGOLD. Our main finding is therefore that in the Mplus implementation, the approximation of the distribution of the LR statistic is much more data dependent than in the LatentGOLD implementation. This effect is stronger (and, thus, the differences between Mplus and LatentGOLD are larger) when the true model is the H0 model than when the true model is the H1 model, and it is also stronger for low class separation than for high class separation. We also found large differences between Mplus and LatentGOLD in the average of the mean and the standard deviation of the estimated distribution of the LR statistic, showing the two implementations derive the p-value of the observed VLMR value from rather different estimated distributions.

Another important finding is that neither of the two implementation yield uniformly distributed p-values under the null hypothesis. The Mplus p-values are closer to uniform than those from LatentGOLD. But, overall, it seems the VLMR statistic is not the best measure for

model selection in mixture models. This aligns with criticism on the VLMR test claiming that the regularity conditions mentioned by Vuong do not hold for mixture models (Jeffries, 2003; Wilson, 2015). It therefore seems better to use the BIC or the BLRT instead. In contrast to the VLMR test, the BLRT does not rely on asymptotic results, but instead constructs the distribution of the likelihood-ratio test statistic of interest by Monte Carlo simulation. Simulation studies by Feng and McCulloch (2003), McLachlan and Peel (1997), and Nylund et al (2007) showed this approach to work well.

We took the simulation setup from Nylund et al (2007) as our starting point since this is the key reference for the comparison of class enumeration measures in LCA and LPA. We selected two somewhat favorable conditions, that is, LCA and LPA with a relatively large sample size, well-separated classes, and equal class proportions. Given the well-separated classes, it was not surprising that the encountered power to reject the model with K-1 classes was 1.00 for these two conditions. In the third condition with bad-separated classes, the Mplus approach showed much larger acceptance rates of the (incorrect) null than the LatentGOLD approach. This is in agreement with what we observed when analyzing the example data set from the UCLA website.

Lo, Mendell, and Rubin (2001) proposed a slightly modified version of the VLMR test, referred to as the adjusted Lo-Mendell-Rubin (aLMR) test. It involves dividing the value of the test statistic (the LR value) by a constant which depends on the sample size and the number of additional parameters when increasing the number of classes by one. For our three simulation conditions, this constant equals 1.016 (LCA-4), 1.016 (LCA-4), and 1.014 (LCA-3). Since the aLMR test uses the same sampling distribution as the VLMR test and since the constant is very close to 1, our results on the Mplus and LatentGOLD comparison also apply to the aLMR test.

As our simulation settings were somewhat limited, future research may involve a more extended comparison between the Mplus and LatentGOLD approach, and may aim to yield a conclusion regarding which method is the one to be preferred. It may also be possible to derive (more extreme) adjustments of the VLMR test yielding more uniformly distributed p-values for a broad range of condition (such as model types, class-separation levels, and sample sizes), in which case the comparative performance of the Mplus and LatentGOLD implementations should be re-evaluated.

Finally, when estimating LCA models, one often obtains boundary solutions. In such cases, Mplus treats the threshold parameters concerned as fixed parameters taking on a large positive or negative value (typically 15 or -15). It is, however, unclear whether this is a valid approach when using the VLMR test. By default, LatentGOLD prevents the occurrence of boundary solutions by using posterior mode estimation; that is, by using Dirichlet priors for the model probabilities. In our simulation, we did not use this option since it unclear whether the VLMR test can be used with posterior mode instead of maximum likelihood estimates of the H0 and H1 models. This is also a topic for future research.

## References

- Feng, Z.D. & McCulloch, C. E. (2003). Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58 (3), 609-617.
- Golden, R. M. (2003). Discrepancy Risk Model Selection Test theory for comparing possibly misspecified or nonnested models. *Psychometrika*, 68(2), 229–249.
- <https://doi.org/10.1007/BF02294799>

- 73 Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables.  
74 *Biometrika*, 48, 419-426.
- 75 Jeffries, N. (2003). A note on “Testing the number of components in a normal mixture.”  
76 *Biometrika*, 90, 991–994.
- 77 Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal  
78 mixture. *Biometrika*, 88, 767-778.
- 79 McLachlan, G. J., & Peel, D. (1997). On a resampling approach to choosing the number of  
80 mixture components in normal mixture models. In L. Billard & N. I Fisher (Eds.),  
81 *Computing science and statistics* (pp. 260-266). Interface Foundation of North America.
- 82 Merkle, E. C., & You, D. (2018). nonnest2: Tests of non-nested models [Computer software  
83 manual]. Retrieved from <http://cran.r-project.org/package=nonnest2> (R package version  
84 0.5-2).
- 85 Merkle, E. C., You, D., & Preacher, K. J. (2016). Testing non-nested structural equation  
86 models. *Psychological Methods*, 21(2), 151–163. doi:10.1037/met0000038.
- 87 Nylund, K.L., Asparouhov, T. & Muthén, B. (2007). Deciding on the number of classes in latent  
88 class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural*  
89 *Equation Modeling*, 14, 535-569.
- 90 Muthén, L.K., & Muthén, B.O. (2002). *Mplus VERSION 2.12: Addendum to the Mplus users’s*  
91 *guide*. Los Angeles, CA: Muthén & Muthén.
- 92 Schneider, L., Chalmers, R.P., Debelak R. & Merkle, E.C. (2020) Model Selection of Nested  
93 and Non-Nested Item Response Models Using Vuong Tests, *Multivariate Behavioral*  
94 *Research*, 55, 664-684, DOI: [10.1080/00273171.2019.1664280](https://doi.org/10.1080/00273171.2019.1664280)

- 95 Vermunt, J. K., & Magidson, J. (2021). *Upgrade manual for Latent GOLD Basic,*  
96 *Advanced/Syntax and Choice 6.0.* Belmont, MA: Statistical Innovations Inc..
- 97 Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses.  
98 *Econometrica*, 57, 307-333.
- 99 Wilson, P. (2015). The misuse of the Vuong test for non-nested models to test for zero-inflation.  
100 *Economics Letters*, 127, 51–53,  
101

## **Appendix: Running the Simulation with the Latent GOLD 6.0 Syntax**

As indicated in the main text, Latent GOLD 6.0 Syntax (Vermunt and Magidson, 2021) was used for the reported simulation study. To generate a data set one needs to define the population model of interest, use the “outfile” option “simulation”, provide a case/frequency weight indicating the sample size, and specify the population parameters as starting values between “{ }” at the end of the equations. For the LPA with 4 classes, the Syntax file “simulate.lgs” contain this the model setup:

```
options
  output parameters standarderrors profile;
  outfile 'sim.txt' simulation;
variables
  caseweight freq1000;
  dependent (y1-y8) continuous;
  latent Cluster nominal 4;
equations
  Cluster <- 1;
  y1 - y8 <- 1 | Cluster;
  y1 - y8;
  {0 0 0
  2 0 0 0
  2 0 0 0
  0 2 0 0
  0 2 0 0
  0 0 2 0
  0 0 2 0
  0 0 0 2
  0 0 0 2
  1 1 1 1 1 1 1 1}
```

The Syntax used to run models from 3 to 5 classes using the generated data file “sim.txt” is as follows:

```
options
  maxthreads all;
```



```
134      startvalues seed=0 sets=32 iterations=250;
135      output parameters standarderrors profile append='LG.csv';
136      variables
137          dependent (y1-y8) continuous;
138          latent Cluster nominal 3:5;
139      equations
140          Cluster <- 1;
141          y1 - y8 <- 1 | Cluster;
142          y1 - y8;
```

143 Note that by requesting models from 3 to 5 classes, one obtains VLMR tests comparing models  
144 with 3 and 4 classes and models with 4 and 5 classes. With “append='LG.csv'”, we indicate that  
145 the a compact version of the output (which includes the VLMR information) should appended to  
146 an output file in csv format. The Mplus version of the VLMR tests are obtained by using  
147 “standarderrors=robust” and adding the keyword “mplus” to the output options. The 1000  
148 replications can be performed by running LatentGOLD in batch mode as follows:

```
149      lg60.exe simulate.lgs estimate.lgs /b /r 1000
```

150 Here the /b switch indicates the program should run in batch model and the /r switch indicates  
151 the models in the specified lgs files should be run multiple times (here 1000 times).

152