ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202105.00042

Random Intercept Latent Transition Analysis (RI-LTA): Disentangling Within-Person Transitions from Between-Person Differences

Authors: Wen Congcong, vermilion, vermilion

Date: 2021-05-07T00:00:00+00:00

Abstract

Traditional latent transition analysis constitutes a single-level analysis, yet it can likewise be conceptualized as a two-level analysis. Muthén and Asparouhov proposed Random Intercept Latent Transition Analysis (RI-LTA) within a single-level analytic framework from a two-level perspective, wherein within-person self-transitions occurring across timepoints can be regarded as analyzed at Level 1, and between-person differences invariant across timepoints can be regarded as analyzed at Level 2, thereby disentangling individuals' self-transitions from initial between-person differences. The random intercept factor f is pivotal in random intercept latent transition analysis. When the factor loadings of observed indicators on the random intercept are substantial, this indicates considerable indicator variance and large initial between-individual differences on that indicator. From this perspective, the random intercept can be viewed as absorbing measurement non-invariance within the model. Precisely because the differences on observed indicators are substantial and discrimination between classes is high, this portion of variance is absorbed by the random intercept without influencing the latent class variable, consequently yielding higher off-diagonal transition probabilities in the latent class transition probability matrix. In conventional latent transition analysis, the high discrimination between classes is not absorbed by a random intercept and directly impacts latent class transition probabilities, resulting in lower off-diagonal transition probabilities and higher diagonal probabilities (probability of remaining in the initial class). The procedure for conducting random intercept latent transition analysis in Mplus software can be summarized in four steps. First, perform latent class analysis (LCA) separately for each timepoint; second, conduct conventional latent transition analysis (LTA) and random intercept latent transition analysis; third, execute Monte Carlo simulation studies utilizing parameter estimates saved from the optimal model obtained in the preceding step; fourth, based on existing theoretical hypotheses or practical research needs, one may incorporate background covariates and distal outcome variables for analysis, or undertake multiple-group analysis, mover-stayer analysis, multilevel random intercept latent transition analysis, or time-varying factor analysis to investigate the data sample at a deeper level. Longitudinal survey data from the 2016 cohort of undergraduates at an elite research university were employed to demonstrate the application of random intercept latent transition analysis. The first three steps of this empirical example aligned with the aforementioned procedure, while the fourth step introduced the covariate "university admission method" for multiple-group and interaction effect analyses. The findings revealed: in the RI-LTA model, students in the high intrinsic and extrinsic motivation group exhibited a 33.0% probability of remaining in their original class, with this group demonstrating greater propensity to transition to other classes; in the conventional LTA model, the same group exhibited a 68.9% probability of remaining, showing greater propensity to stay in the initial class. This illustrates that the random intercept latent transition analysis model avoids overestimating the probability of remaining in one' s original class through the introduction of a random intercept. Future research could employ Monte Carlo simulation studies to investigate the applicability of the random intercept latent transition analysis model, or draw inspiration from multilevel analytic approaches to explore the implementation of multilevel random intercept latent transition analysis in statistical software.

Full Text

Random Intercept Latent Transition Analysis (RI-LTA): Separating Within-Subject Change from Between-Subject Variation

WEN Congcong, International College, Xiamen University, Xiamen 361102, China
ZHU Hong, Graduate School of Education, Peking University, Beijing 100871, China

Abstract

Traditional latent transition analysis (LTA) is typically conducted as a single-level analysis, yet it can also be conceptualized as a two-level analysis. From this two-level perspective, Muthén and Asparouhov (2020) proposed random intercept latent transition analysis (RI-LTA) within a single-level framework. In this approach, within-subject changes across time points are analyzed at Level 1, while time-invariant between-subject differences are analyzed at Level 2, thereby separating self-change from initial between-subject variation and avoiding overestimation of the probability of remaining in the initial class. This study demonstrates the application of RI-LTA using longitudinal survey data from 2016 undergraduate cohorts at an elite research university. The primary advantage of this method is that introducing a random intercept prevents overestimation of transition probabilities for remaining in one' s original group. Future research

could employ Monte Carlo simulation studies to investigate the applicability of RI-LTA and, inspired by multilevel modeling approaches, explore multilevel random intercept latent transition analysis in statistical software.

Keywords: latent transition analysis, random intercept latent transition analysis, single-level analysis, multilevel analysis, Monte Carlo simulation

Introduction

Traditional grouping methods have predominantly focused on known variables such as gender, major, and age. In recent years, person-centered mixture models have gained increasing popularity among researchers (Collins & Lanza, 2009). While variable-centered approaches aim to identify differences between observable subsamples through known grouping variables, person-centered latent class analysis seeks to identify differences between unobservable subsamples (Wang & Wang, 2019).

The progenitor of latent variable mixture models is latent class analysis (LCA), originally proposed by Lazarsfeld in 1950 and further developed by Lazarsfeld, Henry, Goodman, and others. Person-centered clustering methods extend beyond LCA; the traditional K-means clustering method was once considered superior for classifying cases in data samples and continues to be used by many scholars domestically and internationally. However, Magidson and Vermunt (2002) compared these two approaches using simulated data and concluded that LCA is significantly superior to cluster analysis. When research data are longitudinal and contain multiple time points, LCA' s extension—latent transition analysis (LTA)—becomes necessary. Since the 1990s, LTA has gradually entered the 视野 of foreign researchers (Collins & Wugalter, 1992) and has been applied in numerous social and behavioral science studies (Dominic et al., 2017; Jagenow et al., 2015; Pan et al., 2017; Ryoo et al., 2018), medical and public health research (Cochran et al., 2015; Cosden et al., 2015; Kenzik et al., 2015), and educational research (Liu et al., 2016; Ryoo et al., 2015).

Chinese researchers began using LTA in 2011, when Wu Peng and Liu Huashan (2011) first introduced the method at the 90th Anniversary Conference of the Chinese Psychological Society and the 14th National Psychology Academic Conference, demonstrating its application in studying aggressive behavior among middle school students. In 2014, Wu Peng et al. (2014) published a nearly identically titled study in *Psychological Science* investigating three aggression patterns among 276 middle school students in Hubei Province. Subsequently, LTA has been applied to research on self-injurious behavior among Hong Kong middle school students (Hu & Zhang, 2013; Wang et al., 2014; Wang et al., 2015), peer victimization among Jinan middle school students (Yang et al., 2020), developmental trends in adolescent internet addiction (Xie et al., 2018), suicide risk among college students in Hubei (Liu & Ou, 2018), loneliness and its influencing factors among Americans over 50 (Wu et al., 2017; Wu et al., 2018), and depression status and trends among adults across 25 Chinese provinces (Liao &

Lian, 2020). LTA has become a commonly used research method in domestic and international psychology and sociology.

However, controversies remain regarding LTA usage protocols, with inconsistent standards that may lead to researcher subjectivity influencing results. Current debates primarily concern the appropriate procedural flow, which estimation results and statistics should be reported, and how to interpret the validity of final solutions. The LTA model can be expressed as follows (Nylund, 2007; Wang & Wang, 2019), where $\tau_{c_t|c_{(t-1)}}$ represents the probability of an individual transitioning to class $c_t = m$ at time point t given that they belonged to class $c_{(t-1)} = k$ at time $t-1$:

C denotes the number of classes for both c_t and $c_{(t-1)}$, which are generally equal. To regress c_t on $c_{(t-1)}$, the last category of c_t is set as the reference group by default. Since $c_{(t-1)}$ is also a multicategorical independent variable in the model, multinomial logistic regression is required, necessitating the creation of $C-1$ dummy variables $d_1, d_2, \ldots, d_{C-1}$ as indicators for the logits, with d_c also set to 0 as the reference group. α_m is the intercept for transitioning to class m at time point t, while $\beta_{mk} d_k$ represents the log odds ratio of the probability of belonging to class m at time t relative to reference group $c_t = C$, compared to belonging to class k at time $t-1$ relative to reference group $c_{(t-1)} = C$. For each time point $t = 1, 2, \ldots, T$, there are $C-1$ α parameters and $C-1$ β parameters to be estimated. If categorical covariates are to be included in the model, additional terms γ_{mx} and $\gamma_{c_{tx}}$ are required, with the last category of the covariate also set as the reference group.

$$\tau_{c_t|c_{(t-1)}} = P(c_t = m|c_{(t-1)} = k) = \frac{\exp(\alpha_m + \beta_{mk} d_k + \gamma_{mx})}{\sum \exp(\alpha_{c_t} + \beta_{c_{tk}} d_k + \gamma_{c_{tx}})}$$

LTA can be viewed as a single-level analysis examining sample grouping and class transitions across time points. The model implicitly assumes that the correlation of observed indicator variables across time points is fully explained by the correlation between latent class variables. Simultaneously, LTA can also be conceptualized as a two-level analysis, where within-subject changes across time points are analyzed at Level 1 and time-invariant between-subject differences at Level 2 (Muthén & Asparouhov, 2020). Two-level analysis yields more accurate transition probabilities because it separates within-subject change from initial between-subject differences, allowing correlations between latent class variables to explain only within-subject changes across time points.

This concept has been proposed in multiple studies. For instance, research on latent trait-state modeling (Kenny & Zautra, 1995; Cole et al., 2005; Eid et al., 2017) treated invariant between-subject differences as latent trait factors represented by continuous latent variables. Hamaker et al. (2015), using cross-lagged panel modeling as an example, advocated separating invariant between-subject differences (random intercepts) so that cross-lagged relationships where all previous time points only influence their adjacent subsequent time points would not be affected by between-subject differences. Kenny and Zautra (1995) first proposed separating latent trait factors from latent classes. In their model, observed indicators are the sum of latent trait factors, latent classes, and residuals representing measurement error. The crucial feature is that latent trait factors directly influence observed indicators but not latent classes. This specification ensures that latent classes are unaffected by latent trait factors, meaning relationships between latent classes are not influenced by time-invariant between-subject differences.

The common feature of Kenny and Zautra’ s and Hamaker et al.’ s studies is that observed indicators are continuous variables. When both observed indicators and latent variables are categorical, Eid and Langeheine (1999; 2003) examined latent trait-state-latent class modeling, where latent class variables from different time points and latent trait factors jointly influence categorical observed indicators. This model is essentially random intercept latent transition analysis, though it was not named as such at the time. In summary, because conventional LTA does not separate time-invariant between-subject differences, it may produce biased model parameter estimates, particularly substantially biased latent class transition probability estimates. In 2020, Muthén and Asparouhov formally proposed the RI-LTA model, which achieves separation of within-subject change and between-subject differences using two-level analysis thinking within a single-level framework. This model’ s introduction marks a new developmental stage for LTA.

This paper first aims to briefly introduce the RI-LTA model to familiarize domestic researchers with this new method. Second, we demonstrate the RI-LTA process using longitudinal survey data from 2016 undergraduate cohorts at an elite research university. Through this example, we investigate how conventional LTA overestimates transition probabilities, illustrating RI-LTA’ s advantages over traditional methods. Currently, no domestic studies have applied RI-LTA to fit data samples. Additionally, we verify the effectiveness of RI-LTA parameter estimation in this specific context through Monte Carlo simulation of the original model, demonstrating how to test RI-LTA’ s validity in particular situations—an approach not yet employed in domestic or international research. Finally, by introducing an enrollment type covariate, we demonstrate how to conduct multiple-group analysis and interaction effect analysis within the RI-LTA framework, providing references for the method’s application and extension. In summary, this study introduces RI-LTA to China for the first time, aiming to increase awareness and facilitate its use in research. We demonstrate RI-LTA using longitudinal motivation data from undergraduates at an elite research university and provide corresponding Mplus syntax examples in the appendix to achieve this goal.

Research Model Introduction

This section provides a brief introduction to the RI-LTA model to familiarize domestic researchers with this new method.

[Figure 1: see original paper] presents the path diagram of RI-LTA with two time points. Here, f represents the random intercept; $y_{11}, y_{12}, \dots, y_{1p}$ denote p observed indicators at time point t_1; $y_{21}, y_{22}, \dots, y_{2p}$ denote p observed indicators at time point t_2; c_1 represents the latent class variable at time point t_1; and c_2 represents the latent class variable at time point t_2. The random intercept f can be either several continuous latent factors or a multicategorical latent variable. However, based on Muthén and Asparouhov' s (2020) findings from emotion and dating case studies, model fit indices are superior when f is a continuous latent factor rather than a multicategorical latent variable, making continuous latent factors the better choice for random intercepts. To simplify model analysis and reduce unnecessary computation time, the number of random intercept latent factors is set to 1.

Several implicit model assumptions require clarification. The model assumes that between-subject initial differences do not change over time. Factor loadings $\lambda_1, \lambda_2, \dots, \lambda_p$ for the p observed indicators at time point t_1 are set to be exactly equal to their corresponding loadings at time point t_2. To simplify analysis, corresponding factor loadings are also set to be equal across different latent classes for the same indicator. The random intercept f directly influences observed indicators at each time point but does not directly affect latent class variables c_1 and c_2. However, f explains part of the cross-time correlation—that is, between-subject differences—thereby influencing transition probabilities. In RI-LTA with n time points, c_n could potentially influence c_{n-1}, c_{n-2}, or c_{n-3}, but to simplify analysis, the model only assumes that c_n influences c_{n-1}. LTA indicators can be continuous, dummy, multicategorical, ordinal, count variables, etc. However, when the number of observed indicators and response categories is large, the total number of response patterns becomes excessive, leading to some patterns having no or very few respondents and compromising model identification. Therefore, dummy variables with fewer response categories should be used as observed indicators whenever possible.

The random intercept f is crucial in RI-LTA. Large factor loadings of observed indicators on the random intercept indicate high item probabilities and substantial initial differences among cases on those indicators. From this perspective, the random intercept can be viewed as absorbing measurement noninvariance in the model. Because differences on observed indicators are large and discrimination between classes is high, this variation is absorbed by the random intercept without affecting latent class variables, resulting in higher off-diagonal transition probabilities in the latent class transition probability matrix. In conventional LTA, high discrimination between classes is not absorbed by a random intercept and directly influences latent class transition probabilities, yielding lower off-diagonal transition probabilities and higher diagonal transition probabilities (probability of remaining in the initial class). Based on Muthén and Asparouhov' s (2020) empirical examples and Monte Carlo simulation studies, RI-LTA consistently demonstrates superior model fit indices compared to conventional LTA, suggesting that conventional LTA tends to overestimate the probability of remaining in the initial class across different classes.

The RI-LTA model can be expressed mathematically as:

$$P(U_{pit} = 1|C_{it} = j, f_i) = \frac{\exp(\alpha_{pj} + \lambda_p f_i)}{1 + \exp(\alpha_{pj} + \lambda_p f_i)}$$

where U_{pit} is the pth dummy observed indicator for case i at time point t, $C_{it} = j$ indicates that case i at time point t belongs to the jth latent class, α_{pj} represents model parameters for the latent class variable observed indicators, λ_p denotes the factor loading for the pth observed indicator, and f_i represents the factor score for case i, with the factor distribution being $N(0, 1)$. In conventional LTA, the model analyzes the conditional probability of observed indicators given class j, $P(U_{pit} = 1|C_{it} = j)$. In RI-LTA, the conditional probability of observed indicators is conditional on both class j and f_i. When observed indicators are dummy variables, higher positive factor scores on f_i indicate higher probability of responding $U_{pit} = 1$, while higher negative factor scores indicate higher probability of responding $U_{pit} = 0$. Thus, large absolute positive or negative factor scores imply large differences between classes that are unlikely to transition. This demonstrates that if data are generated based on RI-LTA, conventional LTA will tend to overestimate the probability of remaining in the initial class.

Research Steps

RI-LTA can currently only be implemented in Mplus software, requiring the latest version 8.5. The default estimation method is maximum likelihood, capable of analyzing longitudinal data or conducting Monte Carlo simulation studies. The RI-LTA procedure in Mplus can be summarized in four steps:

1. Conduct latent class analysis (LCA) separately for each time point to explore the number of classes contained in the sample at each time point.
2. Conduct conventional LTA and RI-LTA to explore the number of classes and transition probabilities after gradually introducing latent class variables from different time points and the random intercept. Compare model fit indices to identify the optimal model.
3. Use the parameter estimates saved from the previous step to conduct Monte Carlo simulation studies, verifying the validity of parameter estimation for the target model under specific conditions.
4. Based on existing theoretical assumptions or practical research needs, introduce background covariates or distal outcome variables, or conduct multiple-group analysis, mover-stayer analysis, multilevel RI-LTA, or longitudinal factor analysis for deeper investigation of the data sample.

Researchers need not strictly follow all four steps. Steps 1 and 2 are mandatory for RI-LTA, while Steps 3 and 4 are optional. The following section demonstrates RI-LTA using learning motivation data from undergraduates at an elite research university.

Research Example

This study utilized two-wave longitudinal survey data from 2016 undergraduate cohorts at an elite research university, collected during the summer before enrollment through the first month of university (2016) and during the second semester of junior year (2018). The sample comprised 634 cases with complete responses, whose distribution is shown in Table 1 .

The learning motivation scale in the questionnaire included two dimensions: intrinsic and extrinsic motivation. Extrinsic motivation comprised items related to dependence on others' evaluation, interpersonal competition, and pursuit of rewards, while intrinsic motivation included challenge and enthusiasm items. Items used a 4-point scale ranging from "very uncharacteristic" to "very characteristic." In this example, the contingency table for model parameter estimation was derived from original data with 4 response options, 10 observed indicators, and 2 time points, producing 1,099,511,627,776 possible response patterns. The large number of patterns and empty cells in some item options compromised LTA model identification. Additionally, RI-LTA assumes measurement invariance, constraining scores across latent classes in c_1 and c_2 to be equal and factor loadings on random intercept f to be time-invariant. Mismatched class numbers would prevent model convergence and produce biased results. To address these issues, we converted the 4-point observed indicators to dummy variables by scoring "very uncharacteristic" and "somewhat uncharacteristic" as 0 and "somewhat characteristic" and "very characteristic" as 1. This conversion reduced response patterns to 1,048,576 possibilities, effectively avoiding identification problems. The scale demonstrated good reliability and validity with a Cronbach' s alpha of 0.70, meeting statistical analysis standards. The scale structure is presented in Table 2 .

Latent Class Analysis (LCA)

We first conducted LCA separately for 2016 and 2018 using Mplus 8.5, yielding the results shown in Table 3 . For LCA models with different numbers of classes, higher log-likelihood values indicate better model fit, while lower Bayesian Information Criterion (BIC) values indicate better fit. Classes comprising less than 3% of the total sample were considered non-representative (Ryoo et al., 2018). For the 2016 time point, log-likelihood increased with class number, while BIC reached its minimum at 3 classes and increased at 4 classes, indicating a 3-class solution as optimal. For 2018, log-likelihood similarly increased with class number, with BIC minimized at 4 classes and increasing at 5 classes, indicating a 4-class solution as optimal.

Conventional LTA and RI-LTA

We next introduced latent class variables c_1 (2016) and c_2 (2018) simultaneously into conventional LTA and RI-LTA models, producing the results shown in Table 4 . RI-LTA differs from conventional LTA by additionally introducing random intercept f to influence observed variables across all time points, with factor loadings for the same observed variable constrained equal across time points. Overall, when analyzing the same number of classes, RI-LTA estimates additional factor loading parameters equal to the number of observed indicators (10 in this example), yielding relatively larger log-likelihood values and smaller BIC values. This indicates that RI-LTA generally provides superior model fit to conventional LTA, supporting the necessity of introducing a random intercept. The RI-LTA model with $c_1 = 3$, $c_2 = 4$ achieved the maximum log-likelihood, while the model with $c_1 = 3$, $c_2 = 3$ achieved the minimum BIC. Following Muthén and Asparouhov' s (2020) recommendation, when log-likelihood and BIC yield conflicting results, BIC should be prioritized. Additionally, the $c_1 = 3$, $c_2 = 3$ RI-LTA model estimated only 48 parameters, compared to 91 parameters in the $c_1 = 3$, $c_2 = 4$ model, making it more parsimonious. Therefore, the $c_1 = 3$, $c_2 = 3$ RI-LTA solution was selected as optimal.

We then classified and named the three latent classes based on Mplus output. Figure 2 [Figure 2: see original paper] shows response probabilities across items for the three latent classes. All three classes showed high probabilities of endorsing Item 10, while Items 1-9 distinguished the classes. Since Items 1-6 measure extrinsic motivation (dependence on evaluation, interpersonal competition, reward pursuit) and Items 7-10 measure intrinsic motivation (challenge and enthusiasm), we named the classes as follows: Class 1, with relatively low response probabilities on Items 1-9, was named the "Passive Group" ; Class 2, with high probabilities on Items 1-9, was named the "Active Group" ; and Class 3, with low probabilities on Items 1-6 but high probabilities on Items 7-9, was named the "Intrinsic Motivation-Focused Group."

Further examination of results yielded transition probabilities for the three classes across two time points, shown in Table 5 . For Classes 1 and 3, probabilities of remaining in the same class were similar between methods. However, for Class 2, conventional LTA showed a 68.9% probability of remaining in the same class, while RI-LTA reduced this to 33.0%. This demonstrates that random intercept f absorbed initial between-subject differences, and conventional LTA overestimated the probability of Class 2 remaining in its initial class, making RI-LTA the more appropriate method.

Monte Carlo Simulation Study

Using parameter estimates saved from RI-LTA, we generated new datasets for a Monte Carlo simulation study to validate parameter estimation effectiveness under these specific conditions. The model specification maintained 20 dummy observed indicators, 3 classes each for c_1 and c_2, a total sample size of 634, 500 generated datasets, and mixture model analysis type. The inclusion of a random intercept requires additional numerical integration, necessitating corresponding specifications in Mplus analysis commands.

Partial parameter estimation results from the 500 datasets are shown in Table 6

. Selected parameters included the first and tenth factor loadings on the random intercept (λ_1 and λ_{10}), the second intercept for the first class in c_1 (c_1#1 v_2), the tenth intercept for the second class in c_1 (c_1#2 v_{10}), the fourth intercept for the third class in c_1 (c_1#3 v_4), the log odds of belonging to the second class relative to the reference class in c_1 (c_1#2), and the log odds of belonging to the second class relative to the reference class in c_2 (c_2#2). Examination of estimates and biases indicates that RI-LTA accurately estimated parameters under these conditions, with only λ_{10} showing bias of 0.05 (less than 5% of the true value, which is acceptable). Mean square errors were small for all parameters, though λ_{10} was slightly larger at 0.2. The ratio of standard deviation to average standard error fluctuated within the acceptable range of 0.95-1.05 (Asparouhov & Muthén, 2014). Coverage rates for 95% confidence intervals were generally good, except for c_1#2 v_{10}, which showed 0% coverage. These results indicate that most parameters met acceptable criteria for bias, mean square error, standard deviation/average standard error ratio, and coverage, demonstrating good estimation quality. However, the suboptimal estimation for λ_{10} and c_1#2 v_{10} suggests problems with observed indicator y_{10}. Further inspection revealed that all three classes showed high response probabilities on Item 10, with Classes 2 and 3 reaching 1.0 (100% of students responded "characteristic"). This lack of zero responses created a ceiling effect, leading to anomalous parameter estimates—the abnormal intercept values of -996.34 for c_1#2 v_{10} and -286.78 for c_1#3 v_{10}. Nevertheless, such ceiling effects only affect parameter estimates for the relevant observed indicators and do not impact case classification or transition probabilities.

Covariate Analysis: Multiple-Group and Interaction Effects

Following Monte Carlo simulation, researchers may wish to introduce additional background variables for multiple-group analysis or investigate whether these variables affect transition probabilities. This section introduces enrollment type covariates for multiple-group and interaction effect analyses. The enrollment type covariate has three categories: national college entrance exam (NCEE), independent recruitment, and recommendation/special talent. To improve model identification and reduce computational burden, we recoded the covariate into two dummy variables: *dummy*1 (1 = NCEE, 0 = other) and *dummy*2 (1 = independent recruitment, 0 = other). Cases with 0 on both dummies represent recommendation/special talent students.

For multiple-group analysis, we examined both measurement-invariant and measurement-noninvariant models. The measurement-invariant model simply added the two dummy covariates to influence c_1 and c_2 within the RI-LTA framework, while the measurement-noninvariant model additionally specified direct effects of the dummy covariates on observed indicators. Results are shown in Table 7 . Since the measurement-invariant model is nested within the measurement-noninvariant model, chi-square difference testing was appropriate. With 60 additional parameters, the degrees of freedom difference was 60.

Using Satorra and Bentler' s (2010) method, the scaled chi-square difference was 941, yielding $p < 0.001$, rejecting the null hypothesis and favoring the measurement-noninvariant model. However, the measurement-invariant model showed smaller BIC and fewer parameters, making it more parsimonious, so it was used for subsequent interaction effect testing. Examination of results revealed that most regression coefficients between covariates and latent classes were non-significant, except for the effect of $dummy2$ on $c_1\#1$ (-2.364, $p < 0.05$). This indicates that students admitted through independent recruitment had significantly lower log odds of belonging to the Passive Group compared to recommendation/special talent students.

For interaction effect analysis, we examined whether covariates interacted with c_1 to influence c_2. Three models were tested: (1) covariates directly affecting random intercept f only; (2) covariates directly affecting f, c_1, and c_2; and (3) covariates directly affecting f, c_1, and c_2 plus interaction effects between covariates and c_1 on c_2. Results are shown in Table 8 . Since the three models are nested, chi-square difference tests were conducted using Satorra and Bentler' s (2010) method. Both model comparisons yielded $p < 0.001$, favoring the interaction effect model. However, the covariate-affecting-f model showed substantially smaller BIC and fewer parameters, making it the optimal model. Examination revealed that regression coefficients of f on both $dummy1$ and $dummy2$ were non-significant, indicating no significant differences in f across enrollment types. Regression coefficients for $dummy1$ and $dummy2$ influencing c_2 through c_1 were also non-significant, confirming that interaction effects are unnecessary.

Method Summary and Outlook

As described above, RI-LTA represents an advancement over conventional LTA. Its primary contribution is introducing random intercept f to absorb time-invariant between-subject differences, thereby avoiding overestimation of diagonal transition probabilities (remaining in the initial class). Compared to conventional LTA, RI-LTA generally yields better model fit indices, suggesting that most data samples exhibit significant initial between-subject differences. To validate RI-LTA' s parameter estimation effectiveness, researchers should save parameter estimates for Monte Carlo simulation studies. To conduct multiple-group comparisons or investigate interactions with specific background variables, covariates can be introduced for further analysis. This study demonstrated the complete RI-LTA process using longitudinal learning motivation data from 2016 undergraduates at an elite research university, with various model syntax examples provided in the appendix for researchers' reference.

As a newly proposed method in 2020, RI-LTA has numerous derivative models whose effectiveness in empirical research remains to be tested, and many areas require further investigation. For example, when chi-square difference tests, log-likelihood values, and BIC yield conflicting results, model selection criteria need further examination. While using dummy variable indicators facilitates model identification and rapid computation, it may create ceiling or floor effects that produce anomalous parameter estimates—making indicator type selection a worthwhile research question. RI-LTA assumes complete equality of factor loadings for the same observed indicator on random intercept f across time points, implying that factor score increases have identical effects at each time point—an assumption that likely violates real-world data patterns. According to Muthén and Asparouhov' s (2019) reading proficiency example, longitudinal factor analysis models consistently outperformed RI-LTA in fit indices, raising questions about when RI-LTA should be preferred for data interpretation. Muthén and Asparouhov' s (2020) Monte Carlo simulations explored limited conditions; future research should investigate required sample sizes, appropriate numbers of time points, observed indicators, and latent classes for RI-LTA. Furthermore, although RI-LTA is inspired by multilevel modeling thinking, it remains a single-level analysis. Future development will necessarily involve multilevel extensions, though Mplus 8.5 currently cannot execute multilevel RI-LTA—representing a direction for future software development.

References

Hu, J., & Zhang, M. (2013, November). *Latent class and transition analysis of adolescent self-injurious behavior*. In *Psychology and innovation capacity enhancement: Proceedings of the 16th National Psychology Academic Conference* (pp. 374–375). Nanjing, Jiangsu.

Liao, Y., & Lian, R. (2020). Latent transition analysis of adult depressive symptoms: A four-year longitudinal study. *Chinese Journal of Clinical Psychology*, 3, 518–522. doi:10.16128/j.cnki.1005-3611.2020.03.018

Liu, A., & Ou, X. (2018). Category transition of college student suicide risk: A latent transition analysis. *Journal of Southwest University (Social Sciences Edition)*, 2, 104–111+193. doi:10.13718/j.cnki.xdsk.2018.02.012

Wang, B., Zhang, M., Zhang, J., & Hu, J. (2014, October). *Types of adolescent problem behaviors: Results based on latent transition analysis*. In *Abstracts of the 17th National Psychology Academic Conference* (pp. 742–745). Beijing.

Wang, B., Zhang, M., Zhang, J., Hu, J., You, J., & Liang, Y. (2015). Latent transition analysis of adolescent self-injurious behavior: A longitudinal study. *Psychological Science*, 6, 1368–1376. doi:10.16719/j.cnki.1671-6981.20150613

Wu, G., Zhang, M., Ni, Y., Yang, Y., Qi, C., & Wu, J. (2017, November). *Transition and influencing factors of loneliness in older adults: Based on latent transition analysis*. In *Abstracts of the 20th National Psychology Academic Conference—Psychology and National Mental Health* (pp. 722–723). Chongqing.

Wu, G., Zhang, M., Ni, Y., Yang, Y., Qi, C., & Wu, J. (2018). Latent transition analysis of loneliness and its influencing factors in older adults. *Acta Psychologica Sinica*, 9, 1061–1070.

Wu, P., & Liu, H. (2011, October). *Categories and transition of aggressive middle school students: An application of latent class models.* In *Enhancing psychology' s social service consciousness and function: Proceedings of the 90th Anniversary Conference of the Chinese Psychological Society and the 14th National Psychology Academic Conference* (p. 139). Xi' an, Shaanxi.

Wu, P., Liu, H., Chen, J., & Xie, J. (2014). Category transition of aggressive middle school students: Latent transition analysis. *Psychological Science*, 5, 1167–1173. doi:10.16719/j.cnki.1671-6981.2014.05.023

Xie, J., Li, G., Hou, G., & Zhen, F. (2018, November). *Adolescent internet addiction based on LPA: Latent transition analysis.* In *Abstracts of the 21st National Psychology Academic Conference* (pp. 644–646). Beijing.

Yang, X., Chen, L., Cao, Y., Bi, X., Chen, G., & Zhang, W. (2020). Subtypes and stability of early adolescent peer victimization: A person-centered analysis. *Psychological Development and Education*, 2, 146–156. doi:10.16187/j.cnki.issn1001-4918.2020.02.03

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495–508.

Cochran, G., Field, C., & Caetano, R. (2015). Changes in classes of injury-related risks and consequences of risk-level drinking: A latent transition analysis. *The Journal of Behavioral Health Services & Research*, 42(3), 355–373.

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait-state models: Introducing a trait-state-occasion model. *Psychological Methods*, 10(1), 3–20.

Collins, L. M., & Lanza, S. T. (2009). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (Vol. 718). Hoboken, NJ: John Wiley & Sons.

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, 27(1), 131–157.

Cosden, M., Larsen, J. L., Donahue, M. T., & Nylund-Gibson, K. (2015). Trauma symptoms for men and women in substance abuse treatment: A latent transition analysis. *Journal of Substance Abuse Treatment*, 50, 18–25.

Eid, M., Holtmann, J., Santangelo, P., & Ebner-Priemer, U. (2017). On the definition of latent-state-trait models with autoregressive effects. *European Journal of Psychological Assessment.*

Eid, M., & Langeheine, R. (1999). The measurement of consistency and occasion specificity with latent class models: A new model and its application to the measurement of affect. *Psychological Methods*, 4, 100–116.

Eid, M., & Langeheine, R. (2003). Separating stable from variable individuals in longitudinal studies by mixture distribution models. *Measurement: Interdisciplinary Research and Perspectives*, 1, 179-206.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20(1), 102–116.

Jagenow, D., Raufelder, D., & Eid, M. (2015). The development of socio-motivational dependency from early to middle adolescence. *Frontiers in Psychology*, 6, 194.

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, 63, 52–59.

Kenzik, K. M., Martin, M. Y., Fouad, M. N., & Pisu, M. (2015). Health-related quality of life in lung cancer survivors: Latent class and latent transition analysis. *Cancer*, 121(9), 1520–1528.

Liu, Y., Liu, H., & Hau, K. T. (2016). Reading ability development from kindergarten to junior secondary: Latent transition analyses with growth mixture modeling. *Frontiers in Psychology*, 7, 1659.

Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, 20(1), 36–43.

Muthén, B., & Asparouhov, T. (2019). What multilevel modeling can teach us about single-level modeling: Latent transition analysis with random intercepts (RI-LTA). Version 1. Retrieved October 9, 2020, from http://www.statmodel.com/RI-LTA.shtml

Muthén, B., & Asparouhov, T. (2020). Latent transition analysis with random intercepts (RI-LTA). Version 4. Accepted for publication in *Psychological Methods*. Retrieved October 9, 2020, from http://www.statmodel.com/RI-LTA.shtml

Nylund, K. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization* (Unpublished doctoral dissertation). University of California, Los Angeles.

Pan, Y., Liu, H., Lau, P., & Luo, F. (2017). A latent transition analysis of bullying and victimization in Chinese primary school students. *PLOS ONE*, 12(8), e0182802.

Ryoo, J. H., Wang, C., & Swearer, S. M. (2015). Examination of the change in latent statuses in bullying behaviors across time. *School Psychology Quarterly*, 30(1), 105–122.

Ryoo, J. H., Wang, C., Swearer, S. M., Hull, M., & Shi, D. (2018). Longitudinal model building using latent transition analysis: An example using school bullying data. *Frontiers in Psychology*, 9, 675.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, 75, 243–248. doi:10.1007/s11336-009-9135-y

Wang, J., & Wang, X. (2019). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.