



Article scientifique

Article

2025

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

The current practice of latent growth curve modeling in the social and behavioral sciences: Observations and recommendations

Miller, Matt L; Ferrer, Emilio; Ghisletta, Paolo

How to cite

MILLER, Matt L, FERRER, Emilio, GHISLETTA, Paolo. The current practice of latent growth curve modeling in the social and behavioral sciences: Observations and recommendations. In: International journal of behavioral development, 2025, vol. 49, n° 4, p. 389–397. doi: 10.1177/01650254241269723

This publication URL: <https://archive-ouverte.unige.ch/unige:188172>

Publication DOI: [10.1177/01650254241269723](https://doi.org/10.1177/01650254241269723)

The current practice of latent growth curve modeling in the social and behavioral sciences: Observations and recommendations

International Journal of
Behavioral Development
2025, Vol. 49(4) 389–397

© The Author(s) 2024

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01650254241269723

journals.sagepub.com/home/ijbd



Matt L. Miller¹ , Emilio Ferrer² and Paolo Ghisletta³

Abstract

We examine recommendations for three key features of latent growth curve models in the structural equation modeling framework. As a basis for the discussion, we review current practice in the social and behavioral sciences literature as found in 441 reports published in the 19 months beginning in January 2019 and compare our findings to extant recommendations. We then provide suggestions for empirical researchers, reviewing the application of these very popular models, specifically focusing on comparison of alternative change models, time metric and interval features implemented, and the treatment of individually varying time intervals.

Keywords

Latent curve models, longitudinal, guidelines, review

First proposed over 40 years ago (Laird & Ware, 1982), latent growth-curve models (LGCMs) have become one of the most widely used methods for analyzing longitudinal data in behavioral science, particularly in developmental studies (Hershberger, 2003; Preacher, 2019). Their widespread use is due to their ability to model intraindividual change and interindividual differences, to separate measurement error from construct variance, and to provide flexibility for the change function, among other features (Singer & Willett, 2003). Their popularity is increased by their intuitive interpretation given a reasonably brief introduction.

However, the technique's apparent simplicity belies the fact that LGCM is an advanced method (Hoyle, 1995). Like most statistical techniques, LGCMs have underlying assumptions and implementation details that can affect the reliability of their results. Thus, methodologists have published recommendations for their proper use. How often these recommendations are being followed is a legitimate question, given LGCM's popularity. We believe this question should be of interest to two different audiences: first, developmental researchers will find recommendations about better use of LGCM informative in ensuring their research is rigorous and, second, methodologists will learn if their messages are being conveyed effectively.

In this study, we examine current practice of the popular method of LGCM in the structural equation modeling (SEM) framework¹ (McArdle, 1986) to serve both of these goals. We begin by briefly reviewing the LGCM terminology. In the second section, we review common recommendations, for reasons of space, focusing on three important features in the application of LGCM (several other features are reviewed in Supplementary Material; we selected the features for this report based on

shortcomings we found in their application). We then report the results of our review of recent literature in the social and behavioral sciences that applied LGCM. Finally, we discuss the findings of our review and provide recommendations for improving conformance to best practices.

LGCM Terminology

Developmental research is often focused on how people change, involving questions such as how much people change over time, how much individuals vary in change, if there is an association between initial level and amount of change, what the trajectory of change is, or when different groups differ in these characteristics (Baltes & Nesselrode, 1979). The constructs that developmental scientists are interested in measuring are also often considered “latent” (not directly observable) and are thus represented by multiple “manifest” (observable) measurements that indicate the level of that construct, such as test items or performance tasks. Fortunately, LGCM supports both those goals and measurement methods. Readers interested in a detailed review of the basic structure, formulae, and terminology of LGCM will find several

¹Arizona State University, USA

²University of California, Davis, USA

³University of Geneva, Switzerland

Corresponding author:

Matt L. Miller, Department of Psychology, Arizona State University,
900 S. McAllister Avenue, Tempe, AZ 85287, USA.

Email: mlmill47@asu.edu

useful treatments available (Bollen & Curran, 2006; Grimm et al., 2017; Little, 2024; McArdle & Nesselroade, 2014). Here, we will assume a general knowledge of terminology such as slope, intercept, fixed and random effect, residual variance, and fit indices. In the remainder of this section, we review less-familiar terms and introduce concepts and terminology specific to this report.

To help guide this process, we use a hypothetical study as an illustration. Imagine researchers have collected reading skills data on a group of children at five different occasions, intending to measure them at the beginning of third grade and then every 6 months until the beginning of fifth grade. This implies that the “intervals of interest” are 6 months each for a total of five “waves” (data collection occasions). While it may be possible to have multiple collaborators administer reading skills tests to all children at the same time, exactly 6 months apart at each wave, it is much more likely, given a sample of significant size, that participants will vary in the date of data collection, including variations in the interval between waves (both differing between individuals and differing within the same individual; Miller & Ferrer, 2017). This phenomenon is known as sampling-time variation (STV).

A standard way to express an LGCM is through an equation relating outcome variable y to an intercept and a slope. In this expression, time enters the model in a mathematical vector called the “basis vector.” The individual elements of this vector are usually called “basis coefficients” and give the slope its interpretation based on the units of time used for those coefficients. In the top-level formula for the LGCM

$$y_{it} = \alpha_i + \lambda_t \beta_i + \varepsilon_{it}$$

these coefficients appear as the values of λ_t (where y is the measured outcome for individual i at measurement occasion t , α_i is i 's latent intercept, β_i is i 's latent slope, and ε_{it} is i 's residual at occasion t ; we omit the second-level formulae and distributional assumptions for reasons of space). In our example, to set the intercept at the beginning of the study and interpret the slope in linear change per month, the basis would be $[0, 6, 12, 18, 24]'$. A convenient feature of LGCM is that intuitive modifications of the basis coefficients allow accommodation of unequal intervals between waves. For example, if instead of measuring reading skills every 6 months, the schedule had been the beginning of third grade, halfway through third grade, the beginning of fourth grade, the beginning of fifth grade, and the beginning of seventh grade, the researchers could simply change the basis to $[0, 6, 12, 24, 48]'$. The values for the basis vector define the model's metric of time. This metric should reflect the time structure of the actual empirical research.

A strength of the LGCM is its ability to accommodate various different functional forms of change, also known as trajectory shape. The most common are linear trajectories, but many nonlinear shapes are also possible. Perhaps the most powerful feature is LGCM's ability to derive shape from the data instead of pre-specifying the shape. In this case, some of the basis coefficients are parameters to be estimated, changing the basis vector in our example to $[0, \lambda_2, \lambda_3, \lambda_4, 24]'$ with λ_{2-4} to be estimated. Such models are known as “latent-basis models.” A common family of pre-specified, nonlinear shapes are polynomial functional forms, in particular the quadratic shape (Ghisletta et al., 2020). Higher-order polynomials are also possible, leading to cubic models and so forth, though parameter interpretation becomes difficult as

order increases. Loss of degrees of freedom and model convergence can also be challenges. A critical concern in creating higher-order polynomial models is overfitting, leading to loss of generalizability. Other data shapes are available as well, often corresponding to specific, strong hypotheses about the trajectory shape, such as spline models (Cudeck & Klebe, 2002; Ram & Grimm, 2007) and models that are nonlinear in more complex ways (Browne, 1993; Ghisletta et al., 2020; Harring et al., 2020).

Finally, consider the treatment of STV. LGCM in the SEM framework is often thought of as only allowing time to be treated as with equal intervals across all participants. This is not actually a restriction in LGCM, but because it is often assumed, it is frequently handled using “time binning.” This is the common practice of treating individuals as if they had been measured at the same moment even though they were not, thereby discarding STV. For example, if children in our illustration were only measured once per grade, researchers might treat the data as if all children in third grade had been measured at the same moment, even if testing was spread throughout the year. That is, the data of children assessed at the beginning of third grade would be analyzed with the same basis coefficient as those of children assessed at the end of third grade. Time binning is often also observed by age, when participants' ages are rounded to age in years. However, time binning is unnecessary, since LGCM allows individually varying times of measurement by allowing basis coefficients unique to each individual, known as definition variables (Mehta & Neale, 2005; Neale, 1998; Sterba, 2014). This method does not provide many of the familiar fit indices, though model comparisons remain available via likelihood ratio tests (LRTs) and comparison of information criteria (e.g., Akaike information criterion [AIC] or Bayes information criterion [BIC]).

Existing Recommendations for LGCM

Since we have focused this discussion on three features, the recommendations for LGCM that we discuss are hardly exhaustive. General recommendations are available in several excellent texts (Bollen & Curran, 2006; Grimm et al., 2017; Little, 2024; McArdle & Nesselroade, 2014; Singer & Willett, 2003). Here, we enumerate the basic principles and recommendations for reporting the facets that we have selected; specific recommendations for each are provided in the “Discussion.”

Alternative Models of Change

In all SEM analyses, comparing the fit of alternative models is recommended (Kline, 2016; Mueller & Hancock, 2019; Preacher, 2019; Thompson, 2000). In the context of LGCM, this includes comparing different trajectory shapes (McArdle, 1988; Preacher, 2019). The trajectory shape (e.g., no change, linear, quadratic, latent-basis) is important even when it is not the focus of the research (Ram & Grimm, 2007), thus both specification of the basis coefficients and a verbal description of the intended shape are useful (Hesser, 2015; Jackson, 2010; Preacher, 2019). The trajectory shape is always important: misspecified trajectories can lead to bias in parameters of substantive interest and reduce the efficiency of parameter estimation (Mehta & West, 2000; Ram & Grimm, 2007).² Since it is almost never possible to know a priori how estimates would be biased, best practices in determining the most suitable trajectory shape for analysis is required.

An almost-universal recommendation is that the criteria for comparing alternative models and the statistical tests used to compare models be completely reported, which further underlines the general consensus that more than one model should be tested (Hoyle & Isherwood, 2013; Jackson, 2010; Kline, 2016; Mueller & Hancock, 2019; Preacher, 2019).

Metric of Time

How time is conceived and coded—the origin or intercept, the time intervals modeled, the interpretation of all slope and/or shape parameters, the frequency of data collection, and the number of data collection occasions—are central to the specification, estimation, and interpretation of growth models (Collins, 2006; Collins & Graham, 2002; Hopwood et al., 2022; Timmons & Preacher, 2015). Therefore, providing a clear, easy-to-find statement of the triggering event for measurement (that is, what event initiates measurement, such as experiment initiation, birth, and so on), the interval between measurements, and how these fit into the LGCM is essential (Boomsma et al., 2012; Hesser, 2015; Jackson, 2010; Preacher, 2019).

STV and Time Binning

A recognized complication in analyzing longitudinal data in the SEM framework is the treatment of data collection intervals as equal (time binning) despite individual variation in intervals between waves (Sterba, 2014), or STV. Because STV can have important effects (Aydin et al., 2014; Coulombe et al., 2016), some resources recommend providing the mean and standard deviation of data collection intervals (van de Schoot et al., 2017). Explanation and justification of data collection timing and coding of time in the model are also recommended by several resources (Boomsma et al., 2012; Jackson, 2010).

Methods

Because of the large number of studies using LGCM, we limited our literature review to approximately 20 months up to the date of the search. We used the citation search engine Web of Science to search for all records³ with a date between 1 January 2019 and 14 August 2020 that contained “latent growth,” “latent curve,” or “LGCM” in their abstracts, titles, or keywords. We then filtered the resulting list of 646 records to only include disciplines related to psychology, psychiatry, and other social and health sciences to retain our intended scope of review (see Figure 1). We included 39 disciplines and excluded 48 (see Supplementary Material for complete list). We then manually screened titles and abstracts of these records to eliminate methodology studies and other non-empirically focused research; health-related records that did not include a psychological, social, or behavioral component; and a limited number of other records that did not meet our general guidelines for inclusion. Remaining records were reviewed in detail, focusing primarily on the methods and results sections.

Record Selection and Coding

The first step in the detailed review was a final screen for inclusion criteria via direct examination of the record. We required

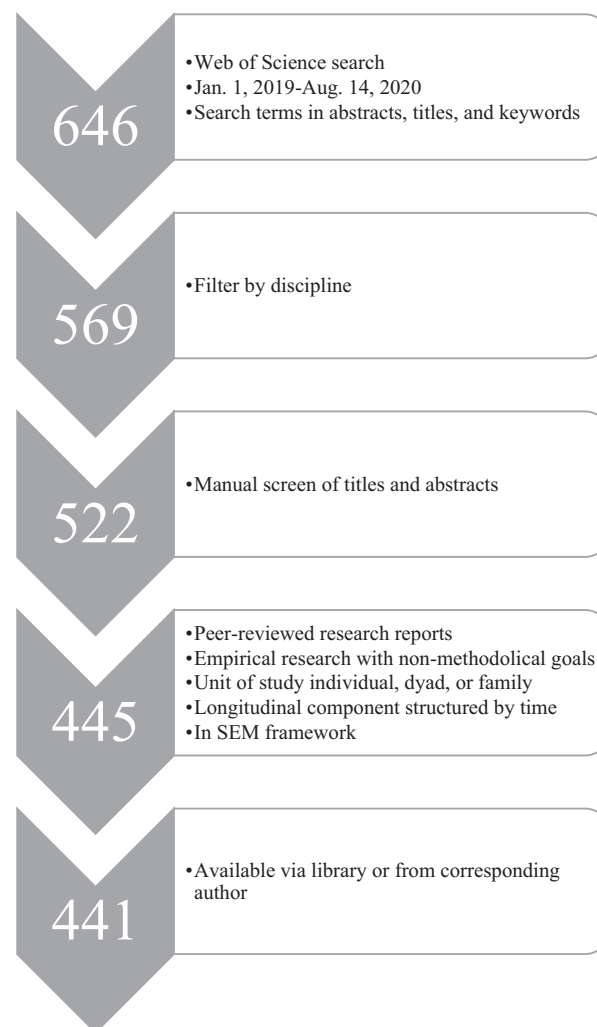


Figure 1. Record Flow Through Inclusion Criteria.

(a) only standard peer-reviewed records (eliminating conference abstracts etc.); (b) empirical research not conducted with a primary aim of establishing or testing methodology; (c) that the unit of study was an individual participant, dyad, or family; (d) that the longitudinal component was primarily structured by metric time (instead of, e.g., after an event regardless of time); and (e) that LGCM was applied in the SEM framework.⁴

Records were accessed online via the library at the University of Geneva, Switzerland. Sixteen records were not accessible due to lack of institutional subscription; in these cases, we emailed the corresponding author. If the author did not respond, we sent follow-up emails three times across two academic semesters. Twelve authors provided materials; four never responded. These latter records were also excluded. Overall, these exclusions resulted in a total of 441 records in the final list for analysis.

Records in the final list were reviewed for how LGCM was applied and reported. When possible, we noted sample size, time metric, number of waves analyzed per unit of analysis, trajectory shape(s) fit to the data, trajectory comparisons, whether details of the timing of data collection were reported, whether longitudinal measurements were treated as occurring simultaneously when

STV was present (time binning), whether corrections were made for time binning, the software used to fit the model(s), and whether fit indices were reported. Only information that was directly available in the record (PDF or web page), including links to supporting materials and Open Science Framework (OSF) page content, were considered. For the 12 records obtained from the corresponding author, only received materials were considered; it is therefore possible that the online version of the article included extra materials that were unavailable to us. If such materials were mentioned in the provided text, we coded it as being present.

Alternative Models of Change

In determining the longitudinal trajectories fit to the data and comparison between the trajectories, we relied primarily on direct description in the record's text. Unfortunately, this was not always sufficient—a substantial number did not directly describe the trajectories used. In such cases, we attempted to determine the functional form by examining path diagrams, parameters constrained to constant values, searching the text for clues such as interpretation of change parameters, and examining the code in the rare cases in which it was provided (21 of the 441 records, just under 5%, provided code).

Metric of Time

Time metric is particularly important in the SEM framework, since model paths are structured based on assumed change across intervals.⁵ Metric falls into two basic groups, continuous time and naturally discrete time. Continuous time includes metrics such as time from an event specific to the participant, time from a common baseline, or age. Naturally discrete time includes grade level in school and day of week. Grade is a case of discrete time, measured from a given baseline, often first grade. On the contrary, "fractional grade" indicates a design in which partial school years are considered; this is considered continuous because change across grade is considered meaningful. A few studies use school year as their time metric; that is, participants in different grades during the same year are combined into one group, but the construct measured is considered only as it changes from year to year, and thus time intervals are considered as naturally discrete. Day of week is another discrete process measured from a day of the week (often Monday) with time structured as discrete intervals to other days.

STV and Time Binning

If records included sufficient information to determine whether data collection intervals were the same for all participants or varied by individual, we coded it as including data-collection timing details. Reporting time interval mean and standard deviation were the minimum required information to be so coded. We also recorded whether or not the issue of time binning was discussed. Naturally discrete metrics of time and data collection methods, which seemed to ensure simultaneous data collection (and thus, strictly equal intervals) were noted. Methods that modeled STV were also recorded. The remaining records, in which STV seemed to be present but unmodeled, were coded as time binned. We note

Table 1. Functional Forms of Change Fit by Presence of Shape Comparisons.

Shape	No comparisons		Comparisons	
No growth	0	(0.0%)	58	(32.0%)
Linear	163	(66.3%)	176	(97.2%)
Quadratic	19	(7.7%)	125	(69.1%)
Cubic	1	(0.4%)	21	(11.6%)
Quartic	0	(0.0%)	1	(0.6%)
Higher-order polynomials	0	(0.0%)	3	(1.7%)
Piecewise linear	8	(3.3%)	7	(3.9%)
Logarithmic	1	(0.4%)	2	(1.1%)
Square-root	1	(0.4%)	0	(0.0%)
Latent-basis	23	(9.3%)	44	(24.3%)
Other	1	(0.4%)	2	(1.1%)
Unspecified	38	(15.4%)	4	(2.2%)

Note. No-comparison studies sum to more than 100% because in a few studies more than one process was modeled using different shapes. Studies with comparisons sum to more than 100% because each study necessarily had at least two shapes.

that in many cases the absence of detailed information about the timing of data collection made this coding somewhat uncertain.

For cases in which time binning was not present (for example, the time metric was naturally discrete), records mentioning that their methods eliminate issues with time binning were coded affirmatively. In most cases, the description of methods made it clear when individuals varied in data collection times. However, there were many cases in which the evidence was ambiguous. When simultaneity was clearly tractable, such as when there was a single school and missing data were reported as due to class absence, we coded the record as having only simultaneous data collection. In cases in which simultaneity was clearly impossible, such as when a range of collection dates was provided, the record was marked as time binned (when the metric of time was continuous). Similarly, we assumed simultaneous data collection when automated devices were used or when online questionnaires were accessible for brief time frames. Other cases were marked as uncertain as to time binning. In general, we attempted to be conservative by assuming that researchers had taken as many precautions as could reasonably be expected to ensure equal time intervals.

Results

Alternative Models of Change

Most studies did not report functional forms of trajectories, fitting just one shape to their data ($n=246$, 55.8%); in two of these cases, only a linear trajectory was possible because the data only included two time points in a non-accelerated design. In 14 studies (3.2%), it was unclear whether shape comparisons had been made (see Table 1). The remaining records (41.0%) reported two or more functional forms compared to determine the best-fitting model. Linear trajectories were the most common across both studies without and with comparisons. The next most common trajectory in studies without comparisons were latent-basis models; in studies with comparisons, quadratic was the second most common. The number of studies that explicitly stated the shape

Table 2. Time Metric Applied.

Metric of time	Count	Percentage of category	
Continuous time	406	92.1	
Age	—	73	18.0
Fractional grade	—	25	6.2
Other trigger	—	308	75.9
Discrete time	34	7.7	
Grade	—	25	73.5
School year	—	2	5.9
Day of week	—	2	5.9
Other	—	5	14.7
Unspecified	1	0.2	

Note. Subcategory percentages are proportion of parent category.

of the time course is lower than reported, since it was occasionally possible to determine the functional form from clues such as slope loadings in figures. There were also cases in which the functional form was specified but was well-hidden in places outside the main text, such as in supplemental materials, generally along the lines of a description or title including “linear slope.”

When functional form was compared ($n=179$), it was most often between linear and quadratic (36.3%). The next most common comparison was between linear and latent-basis (10.1%); no-growth and linear (10.1%); and between no-growth, linear, and quadratic (9.5%). Less common comparisons were between linear, quadratic, and latent-basis (7.3%); no-growth, linear, and latent-basis (6.1%); no-growth, linear, quadratic, and cubic (5.6%); and linear, quadratic, and cubic (3.9%). All other comparisons occurred in fewer than five studies each.

Metric of Time

Almost all studies used a continuous metric of time to structure the LGCM (see Table 2). A small proportion of studies was structured by a naturally discrete metric of time; of these, the great majority (and 6% of the overall sample) were structured by grade in school. Eight studies based on a discrete metric of time (24%) had some substantial conflation of time with a continuous measure of time, for example, by setting up the hypotheses and analysis by grade in school but including conclusions and inferences in terms of continuous time.

STV and Time Binning

Few records reported the details of data collection timing in enough detail that the presence and degree of inter-participant timing variation could be assessed. Only 53 records (12.0%) reported these details to even a marginal degree. Even fewer records ($n=19$, 4.3%) mentioned time binning. Even when considering only studies in which time binning might be an issue, that is, when neither the time metric was naturally discrete nor data were collected simultaneously, only 4% (15/378) discussed time binning. This unconcern for temporal imprecision and its possible effects is further reflected in the handling of time. Of the 378 studies in which time binning was present, less than 5% seem to have applied a method to correct for STV (see Table 3).

Table 3. Time Binning and Corrections.

Handling of time	Count	Percentage of category		
Binning possible	378	87.9		
No correction applied	—	361	95.5	
Correction applied	—	17	4.5	
TSCORES	—	—	11	64.7
Other definition variable approach	—	—	2	11.8
Other correction	—	—	4	23.5
Binning not possible	52	12.1		
Naturally discrete metric of time	—	34	65.4	
Only simultaneous measurements used	—	18	34.6	

Note. Subcategory percentages are proportion of parent category. Eleven records for which the handling of sampling-time variation was unclear are omitted.

Discussion

The three features that we chose to review have mediocre to poor compliance to recommendations. In particular, features related to the metric of time either have gaps in reporting or are not implemented following guidelines. It is possible that some of these omissions were not reported in the records but were actually applied. In such cases, the authors may have omitted reporting in an attempt to meet article length guidelines or as a result of the reviewing process. Thus, we hope any suggestions we offer may be useful to both authors and editors of publications that include LGCM-based research.

Reporting of functional form of change was adhered to in a large majority of cases, but with enough omissions to cause concern. A few of the records omitting functional form were studies in which more than one trajectory was compared for best fit, perhaps with the hazardous idea that rejected shapes were not substantively important, but even among records only fitting a single trajectory, a significant percentage did not mention its functional form. There were a substantial number of studies for which no small effort was required to determine the trajectory. This is a key feature that should be easily located in reported results.

Two recommendations were poorly adhered to across much of the reviewed literature: comparison of multiple models of change and handling of STV. More than half of the studies reviewed reported fitting only one data shape, despite wide evidence that LGCM trajectories cannot be confirmed by fit index threshold and, instead, trajectory shapes should be compared (Newsom, 2015; Ram & Grimm, 2007). Several articles asserted finding evidence for linear growth in the absence of comparison conditions. In multiple studies, when functional forms were compared, researchers proceeded from most to least parsimonious (e.g., from no-growth to linear to quadratic), stopping when fit indices passed some threshold without testing further functional forms. That is, additional trajectories were only tested when model fit was deemed to be “bad.” This is not recommended for testing trajectory shape, since it is quite possible for models to produce fit indices within even conservative guidelines despite being specified with the wrong function (Miller & Ferrer, 2017; Newsom, 2015). Rigorous methodology recommends testing multiple functional forms, including more complex trajectories (Preacher, 2019). Failure to compare trajectories was often

justified by assertions that only linear change could be modeled with three time points, which is untrue. Even with only two measurement occasions,⁶ a no-growth model could be compared to a linear change model. Latent-basis models would also be a reasonable alternative for three or more occasions.

Less than an eighth of the records reported any details on the timing of data collection, though a great majority of the reviewed records likely included STV. In many cases, methods that manifestly required significant time windows for data collection (for example, studies with hundreds of participants requiring multi-hour home visits) made no mention whatsoever of data-collection timing. Although some studies may only be interested in measuring change without regard to time as a metric, if there is population consistency in change per unit time, the possibility of decreased statistical efficiency and biased estimates should still be of concern (Estrada & Ferrer, 2019; Mehta & West, 2000). Techniques exist to account for this variation (Grimm et al., 2017; Sterba, 2014), but very few records used any such technique. This would be less worrisome if the remaining studies had acknowledged the unaccounted-for time binning and discussed assurances against bias and loss of efficiency. However, of those 361 studies, only three studies mentioned time binning.

Although general adherence to many basic methodological recommendations was high in the reviewed records, many had omissions and adherence was concerningly low for the features we have focused. In addition, in our review, we noted a number of statements in records indicating problems with the application of statistical analysis and interpretation, such as the meaning and use of random effects. These and other methodological issues cannot be laid solely at the feet of the authors. Although it is incumbent on researchers to understand the methods they use in enough detail to perform analyses without error in application or interpretation, it is also the responsibility (some might say the prime purpose) of peer-reviewed journals to provide expert review of articles to correct, and in the most egregious cases to reject, those with errors or important omissions.

Suggestions for Implementing Best Practices

We reiterate these suggestions in the hope that better adherence will improve research practice. We recognize that there are sometimes valid reasons for deviation from best practices, but suggest that when they occur, these reasons should be reported and properly justified.

Alternative Models of Change. Alternative forms should be compared using prespecified, justifiable model comparison approaches to determine the trajectory as closely as possible (Newsom, 2015). Fit indices for a single model alone, no matter how good, are insufficient to determine functional form of change; at least one comparison must be provided to justify the functional form (Grimm et al., 2017; Miller & Ferrer, 2017; Preacher, 2019). Chi-square LRTs are often used for comparisons between nested models (Kline, 2016). Following usual practice, the authors should directly report these statistics. All models of change have at minimum two alternative functional forms that can be compared: no-growth and linear change (Wu et al., 2009). The no-growth model or intercept-only model, in which there is no slope term and thus individuals are modeled as having a constant level

of the measured construct, not only can serve as a comparison model when other models cannot but, when the variance of the intercept is set to zero, can also serve as the proper baseline model for fit indices that require a comparison model (Widaman & Thompson, 2003).⁷ Models of change with three or more occasions can additionally be compared to a latent-basis model (or another, non-latent-basis, manually specified model, though this requires strong a priori theory of shape). A latent-basis model will almost always provide the best fit but will absorb degrees of freedom, and may thus be compared for significant improvement in fit via LRTs (Bollen & Curran, 2006; Kline, 2016) or by examining fit measures that impose a penalty for excessive parameters, such as the Bayesian Information Criterion (Burnham & Anderson, 2002; Wagenmakers & Farrell, 2004).

Many other functional forms are possible, including polynomials, splines with fixed knot points, and other change-point models with estimated knot points (Cudeck & Harring, 2007; Cudeck & Klebe, 2002; Grimm et al., 2017). Researchers should compare several models of change, guided by theory and by examination of the data, though the number of measurement occasions will constrain available functional forms (Bollen & Curran, 2006). Processes known to have acceleration and longitudinal data with a clear curvilinear trend should have some non-linear functional form compared, again considering constraints imposed by the number of time points (Ghisletta et al., 2020). It is recommended that all functional forms be clearly reported, including whether the form is based on a priori theory or on examination of the current sample. Besides providing comparison of different functional forms, researchers should also consider justifying likely forms (from prior research, existing theory, or characteristics of the current data) that have not been tested. The tests used to establish the best fitting model should also be fully and clearly reported, including complete reporting of resulting statistics.

Metric of Time. The metric of time, the model's origin of time, and time intervals are essential to any longitudinal study, and are certainly key to LGCM. Therefore, in planning research, these should be carefully considered. It is best to select an analytical metric as close to the natural metric of the phenomenon under inquiry as possible. The guiding principle should be the selection of a metric that accords with hypotheses and the standards of the discipline in which the research is conducted; consideration should be given to the ability to justify any unusual metric on theoretical grounds instead of convenience. We generally recommend the selection of continuous metrics,⁸ simply because they need less justification; researchers should usually justify the use of discrete metrics since important information can be obscured by discarding information on timing (though cases of discrete change do exist; see, for example, Lampl, 1993).

The model's intercept should be chosen based on interpretability, especially when interactions are present, though statistical considerations may sometimes dictate its choice (e.g., the scaling of the basis coefficients, Rovine & Molenaar, 1998). Time intervals are often a function of resources available for data collection: how long the study can run and how many data collection occasions can be afforded. Nonetheless, within these constraints, intervals should be carefully thought out based on hypothesized functional form of change (Biesanz et al., 2004) and research

goals. Whatever origin and however many intervals are chosen, the model itself should strictly reflect these decisions in the selected metric (Bollen & Curran, 2006). Multiples of the time metric are also reasonable when computationally necessary, for example, when change in years is of small magnitude. In such cases, changing the basis coefficients to reflect change in decades can bring the variance of change into the same order of magnitude as other components of the model, facilitating estimation. Latent-basis models could, of course, have many basis coefficients freely estimated, but the anchor points (the fixed basis coefficients require to set scale) should follow this guidance, and the estimated coefficients should be interpreted in light of the actual data collection intervals.

Another consideration is when data collection occurs in waves instead of at a single time point per measurement occasion—that is, when STV is present. A clear, easy-to-find statement of the triggering event for measurement, the interval between measurements, and how these fit into the LGCM is essential in published research, given their importance in the meaning of the model. Whatever decisions are made, all of these dimensions should be clearly stated in the research report, along with justifications for the final decisions.

STV and Time Binning. Enough information should be present for both peer reviewers and readers to evaluate the possible influence of time binning on the model. Researchers should consider the effects of variation in intervals in their studies and report those considerations and any methods taken to account for them. When time binning is present, time should only be coded at the temporal center of data collection waves, that is, as closely as reasonable to the midpoint between the beginning and end of each wave, especially at the end points of the study, to prevent bias due to unbalanced variation between participants (Estrada & Ferrer, 2019; Miller & Ferrer, 2017). This improves the likelihood that time binning effects will be ignorable, particularly when the variation within bins is small in comparison to the interval between bins. In other cases, time binning may lead to parameter bias or loss of statistical efficiency (Aydin et al., 2014; Coulombe et al., 2016). Techniques exist to model individual variation in time intervals, such as the TSCORE option in Mplus (Muthén & Muthén, 2017) or more general definition variable approaches in OpenMx or other suitable software (Grimm et al., 2017; McNeish & Matta, 2020; Sterba, 2014). If there is substantial STV and one of the methods that account for it is not applied, the researcher should justify the decision to use time binning instead of a definition-variable approach. We also suggest that providing at least a basic understanding of how individual participants may have varied in time intervals would also be useful to provide insight into the efficiency and reliability of parameter estimates (van de Schoot et al., 2017). In all cases, the research report should include sufficient information to allow for replication.

Editorial Process Recommendations. Finally, we consider another recommendation external to the work of empirical researchers, rather than being in the hands of editors and reviewers. We believe that editors should encourage all reviewers for LGCM-related articles to pay close, well-informed attention to the analyses. This is certainly the goal of most editors and reviewers, but given some of the oversights observed, in some cases, we

feel that this goal was not achieved. More generally, peer review would benefit from careful review of quantitative methods, examining them as part of the substance of the review instead of as a secondary matter, particularly for more complex techniques. Reviewers and editors should also consider fortifying their evaluations of statistical methodology by comparing articles with guidelines for review, for example, as provided by Hancock and colleagues (2019) in their comprehensive text for reviewers. Of course, authors can also use such resources to evaluate how closely their methods and reporting have adhered to best practices; we recommend such self-review to researchers both when planning studies and when writing up results. Methodological and reporting rigor has already been improved in various domains based on similar recommendations (Appelbaum et al., 2018; Wilkinson et al., 1999).

Conclusion

LGCM is a widely used technique for the analysis of longitudinal data in the social and behavioral sciences. Its intuitive interpretation and the relative ease with which it can be applied given appropriate software belie its underlying complexity, which has led many methodologists to provide recommendations for its use. Focusing on three specific features, we have seen that some recommendations are overlooked in a concerningly high proportion of cases. We have reiterated some suggestions reviewing best practices that we hope are also helpful by being flexible, including a suggestion for more diligent assessment of methodology in the editorial and peer reviewer process. We hope that this article will inspire researchers to carefully consider these key features so as to improve LGCM practice in future research.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the European Commission, Horizon2020 under grant agreement number: 732592-Lifebrain-H2020-SC1-2016-2017/H2020-SC1-2016-RTD and by NIH grant number T32A039772, “Research training in drug abuse prevention: Closing the research-practice gap.”

ORCID iD

Matt L. Miller  <https://orcid.org/0000-0002-2377-6325>

Supplemental material

Supplemental material for this article is available online.

Notes

1. LGCMs in the SEM and multilevel modeling (MLM) frameworks are largely equivalent when time is coded identically (Curran, 2003; McArdle & Hamagami, 1996), but to focus our discussion, we examine LGCM in SEM. Herein, unless a distinction is required, we use LGCM as an abbreviation for LGCM in the SEM framework. Our recommendations apply equally to LGCM in the MLM framework.
2. In our discussion of trajectory shape, we consider only LGCMs with the same model of change applied uniformly to all units of analysis. Other analyses allowing for different

models for latent groups within the data, such as growth mixture models, are areas of active research (Bauer, 2007), useful guidelines for which have been proposed by Bauer and by van de Schoot et al. (2017).

3. Web of Science searches did not include most book chapters at the time this study was conducted.
4. Many records did not specify the analysis framework. However, this could often be inferred by software; for example, the R package lavaan only provides SEM (or multilevel SEM, but not multilevel analyses alone). In other cases, it could be inferred from fit indices reported that are unique to SEM, for example, RMSEA or CFI. Occasionally, we used ambiguous indications, such as figures representing an SEM realization. In more ambiguous cases, we coded the article as using an unspecified framework and thus excluded it.
5. This does not apply to dynamic SEM models, in which instantaneous change across infinitesimal time is modeled, but such models remain extremely rare in current usage (only one in the literature reviewed herein). However, these models also require the correct specification of time intervals in the observed data.
6. In an accelerated, or cohort-sequential, design, more complex trajectories are possible even when individuals have only two measurement occasions; see, for example, the study by Estrada & Ferrer (2019).
7. Note that the default comparison model (also known as the baseline model) in most software is not, in fact the no-growth, fixed-intercept model, so these must often be specified directly by the researcher to obtain meaningful values for popular fit indices such as the Tucker–Lewis Index (TLI) and the Comparative Fit Index (CFI).
8. Note that we refer simply to whether or not time of sampling is rounded. We are not suggesting that studies must use continuous time in a dynamic model (Voelkle et al., 2018), since static models are still appropriate for many questions in the social sciences.

References

- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Aydin, B., Leite, W. L., & Algina, J. (2014). The consequences of ignoring variability in measurement occasions within data collection waves in latent growth models. *Multivariate Behavioral Research, 49*, 149–160. <https://doi.org/10.1080/00273171.2014.887901>
- Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). Academic Press.
- Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*, 757–786. <https://doi.org/10.1080/00273170701710338>
- Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*, 30–52. <https://doi.org/10.1037/1082-989X.9.1.30>
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Wiley.
- Boomsma, A., Hoyle, R. H., & Panter, A. T. (2012). The structural equation modeling research report. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 341–358). Guilford Press.
- Browne, M. W. (1993). Structured latent curve models. In C. M. Cuadras & C. R. Rao (Eds.), *Multivariate analysis: Future directions 2* (pp. 171–197). Elsevier.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach* (2nd ed.). Springer.
- Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology, 57*, 505–528. <https://doi.org/10.1146/annurev.psych.57.102904.190146>
- Collins, L. M., & Graham, J. W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: Temporal design considerations. *Drug and Alcohol Dependence, 68*(s1), 85–96. [https://doi.org/10.1016/S0376-8716\(02\)00217-X](https://doi.org/10.1016/S0376-8716(02)00217-X)
- Coulombe, P., Selig, J. P., & Delaney, H. D. (2016). Ignoring individual differences in times of assessment in growth curve modeling. *International Journal of Behavioral Development, 40*, 78–86. <https://doi.org/10.1177/0165025415577684>
- Cudeck, R., & Harring, J. R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology, 58*, 615–637. <https://doi.org/10.1146/annurev.psych.58.110405.085520>
- Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed effects models for repeated measures data. *Psychological Methods, 7*, 41–63. <https://doi.org/10.1037/1082-989X.7.1.41>
- Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*, 529–569. https://doi.org/10.1207/s15327906mbr3804_5
- Estrada, E., & Ferrer, E. (2019). Studying developmental processes in accelerated cohort-sequential designs with discrete- and continuous-time latent change score models. *Psychological Methods, 24*, 708–734. <https://doi.org/10.1037/met0000215>
- Ghisletta, P., Mason, F., von Oertzen, T., Hertzog, C., Nilsson, L., & Lindenberger, U. (2020). On the use of growth models to study normal cognitive aging. *International Journal of Behavioral Development, 44*, 88–96. <https://doi.org/10.1177/0165025419851576>
- Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Press.
- Hancock, G. R., Stapleton, L. M., & Mueller, R. O. (Eds.). (2019). *The reviewer's guide to quantitative methods in the social sciences* (2nd ed.). Routledge.
- Harring, J. R., Strazzeri, M. M., & Blozis, S. A. (2020). Piecewise latent growth models: Beyond modeling linear-linear processes. *Behavior Research Methods, 53*, 593–608. <https://doi.org/10.3758/s13428-020-01420-5>
- Hershberger, S. L. (2003). The growth of structural equation modeling: 1994–2001. *Structural Equation Modeling: An Interdisciplinary Journal, 10*(1), 35–46. https://doi.org/10.1207/S15328007SEM1001_2
- Hesser, H. (2015). Modeling individual differences in randomized experiments using growth models: Recommendations for design, statistical analysis and reporting of results of internet interventions. *Internet Interventions, 2*, 110–120. <https://doi.org/10.1016/j.invent.2015.02.003>
- Hopwood, C. J., Bleidorn, W., & Wright, A. G. (2022). Connecting theory to methods in longitudinal research. *Perspectives on*

- Psychological Science*, 17, 884–894. <https://doi.org/10.1177/17456916211008407>
- Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*. SAGE.
- Hoyle, R. H., & Isherwood, J. C. (2013). Reporting results from structural equation modeling analyses in Archives of Scientific Psychology. *Archives of Scientific Psychology*, 1, 14–22. <https://doi.org/10.1037/arc0000004>
- Jackson, D. L. (2010). Reporting results of latent growth modeling and multilevel psychology analyses: Some recommendations for rehabilitation psychology. *Rehabilitation Psychology*, 55(3), 272–285. <https://doi.org/10.1037/a0020462>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- Laird, N. M., & Ware, A. H. (1982). Random-effects models for longitudinal models. *Biometrics*, 38, 963–974. <https://doi.org/10.2307/2529876>
- Lamp, M. (1993). Evidence of saltatory growth in infancy. *American Journal of Human Biology*, 5(6), 641–652. <https://doi.org/10.1002/ajhb.1310050607>
- Little, T. D. (2024). *Longitudinal structural equation modeling* (2nd ed.). Guilford Press.
- McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, 16, 163–200. <https://doi.org/10.1007/BF01065485>
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. B. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). Plenum Press.
- McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 89–124). Psychology Press.
- McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. American Psychological Association.
- McNeish, D., & Matta, T. H. (2020). Flexible treatment of time-varying covariates with time unstructured data. *Structural Equation Modeling: A Multidisciplinary Journal*, 27, 298–317. <https://doi.org/10.1080/10705511.2019.1627213>
- Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, 10, 259–284. <https://doi.org/10.1037/1082-989X.10.3.259>
- Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, 5, 23–43.
- Miller, M. L., & Ferrer, E. (2017). The effect of sampling-time variation on latent growth curve models. *Structural Equation Modeling: A Multidisciplinary Journal*, 24, 831–854.
- Mueller, R. O., & Hancock, G. R. (2019). Structural equation modeling. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd ed., pp. 445–456). Routledge.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.).
- Neale, M. C. (1998). Modeling interaction and nonlinear effects with Mx: A general approach. In G. Marcoulides & R. Schumacker (Eds.), *Interaction and non-linear effects in structural equation modeling* (pp. 43–61). Erlbaum.
- Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge.
- Preacher, K. J. (2019). Latent growth curve models. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd ed., pp. 178–192). Routledge.
- Ram, N., & Grimm, K. J. (2007). Using simple and complex growth models to articulate developmental change: Matching method to theory. *International Journal of Behavioral Development*, 31, 303–316. <https://doi.org/10.1177/0165025407077751>
- Rovine, M. J., & Molenaar, P. C. (1998). The covariance between level and shape in the latent growth curve model with estimated basis vector coefficients. *Methods of Psychological Research Online*, 3(2), 95–107. <https://www.dgps.de/fachgruppen/methoden/mpr-online/issue5/art7/article.html>
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford University Press.
- Sterba, S. K. (2014). Fitting nonlinear latent growth curve models with individually varying time points. *Structural Equation Modeling: A Multidisciplinary Journal*, 21, 630–647. <https://doi.org/10.1080/10705511.2014.919828>
- Thompson, B. (2000). Ten commandments of structural equation modeling. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 261–283). American Psychological Association.
- Timmons, A. C., & Preacher, K. J. (2015). The importance of temporal design: How do measurement intervals affect the accuracy and efficiency of parameter estimates in longitudinal research? *Multivariate Behavioral Research*, 50(1), 41–55. <https://doi.org/10.1080/00273171.2014.961056>
- van de Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-Checklist: Guidelines for reporting on latent trajectory studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(3), 451–467. <https://doi.org/10.1080/10705511.2016.1247646>
- Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2018). The role of time in the quest for understanding psychological mechanisms. *Multivariate Behavioral Research*, 53, 782–805.
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11, 192–196. <https://doi.org/10.3758/BF03206482>
- Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, 8(1), 16–37. <https://doi.org/10.1037/1082-989X.8.1.16>
- Wilkinson, L., Task Force on Statistical Inference, American Psychological Association, & Science Directorate. (1999). Statistical methods in psychological journals: Guidelines and explanations. *American Psychologist*, 54(8), 594–604. <https://doi.org/10.1037/0003-066X.54.8.594>
- Wu, W., West, S. G., & Taylor, A. B. (2009). Evaluating model fit for growth curve models: Integration of fit indices from SEM and MLM frameworks. *Psychological Methods*, 14(3), 183–201. <https://doi.org/10.1037/a0015858>