Research Article

Latent Growth Curve Modeling of Ordinal Scales: A Comparison of Three Strategies

Chongming Yang[1]*, Joseph A Olsen[1], Sarah Coyne[2] and Jing Yu[3]

[1]Brigham Young University, Provo, Utah, USA
[2]National Institute of Child Health and Human Development, USA
[3]University of Maryland, Baltimore County, USA

Abstract

Ordinal scales can be used in latent growth curve modeling in three ways: mean, weighted mean scores, and factors measured by scale items. Sum and mean scores are commonly used in growth curve modeling in spite of certain discouragement. It was unclear how much bias these practices could produce in terms of the change rates and patterns. This study compared three methods with Monte Carlo Simulations under different number of response categories of the items, in terms of five key parameters of growth curve modeling. The hypothetical population models were derived from real empirical data to generate datasets of binary, trichotomous, five- and seven-point scales with sample size of 300. Latent growth curve modeling of mean, weighted mean, and factors measured by the ordinal scales were respectively fit to these datasets. Results indicated that modeling the factors that are measured with ordinal scales yield the fewest biases. Biases of modeling the means and weighted of the scales were under one decimal point in the change rates, whereas biases in the variances and covariance of the intercept and slope factors were large. In conclusion, it is inadvisable to use means or weighted means of ordinal scales for latent growth curve modeling. It produces the best results modeling the factors that are measured with the ordinal scales.

Keywords: Monte Carlo simulation; Latent growth curve modeling

Introduction

Ordinal scales that consist of ordered sets of categorical response options are widely used in research and are not true measures of psychological traits or states, which are supposed to be normally distributed random variables. Researchers have recommended that ordinal scales not be used directly as true measures of latent psychological traits or states, alternatively referred to as latent variables, latent constructs, or factors. "In strictest propriety the ordinary statistics involving means and standard deviation ought not to be used with these scales, for these statistics imply a knowledge of something more than the rank order of the data" [1]. Hayes [2] commented that "the problem of measurement, and especially attaining interval levels scales, is an extremely serious one for social and behavioral sciences. It is unfortunate that in their search for quantitative methods researchers sometimes overlook the question of level of measurement…" Treating ordinal scales as continuous data in statistical modeling would produce biased estimates [3-5]. It is common to use the sum or mean scores of scales items for latent growth curve modeling that particularly involves mean and variance of a change. It is not clear to what extent such practice could bias the change and variance estimates in latent growth curve modeling, as compared to an appropriate procedure.

Ordinal indicators reflect latent variables best through probability models [6]. The original items can be specified to measure latent variables through various measurement models in growth curve modeling. In contrast, the means of sets of ordinal items cannot be directly equated to the latent variables. Figure 1 below illustrates an appropriate way to estimate change in a repeatedly measured factors with ordinal indicators [7]. In this "curve of factors" multiple-equation growth curve model, three ordinal items (labeled as Y) can be linked to first-order factors at each of four time-points via probit or logistic factor loadings. Details of the equations are provided by Muthén and Shedden [8].

Some parameters are critical in the latent growth curve modeling. The estimated initial level and change over time are captured by two

second-order latent variables, namely, the intercept and slope. The variances of the intercept and slope factors indicate the individual differences in their initial levels and change rates. The covariance of the two factors indicates the extent to which the initial level is associated with the change rate. The fixed loadings for the slope factor serve to scale the time variable, which is alternatively referred to as time scores. A logarithmic curve pattern can be estimated by specifying the slope factor effects on the repeatedly measured factors to be 0, 0.69, 1.10, and 1.39, which are respectively the natural logs of 1, 2, 3, and 4 (a linear pattern). Different patterns can be estimated by changing the times scores. A model with the best pattern can be selected through model comparisons in terms of smallest Bayesian information criteria or χ^2 of model fit.

Using sums or means of the observed Y variables for each factor in Figure 1 reduces the size of the model such that the four factors are replaced by four observed variables, namely, the sum or mean of scales. Technically, when scale sums or means are used, multivariate normality is assumed and the variables are treated as continuous measures, often using maximum likelihood estimation. However, the multivariate normality assumption is usually violated, resulting in potentially biased estimation of the structural parameters. When the original ordinal observed variables in Figure 1 are specified as categorical, a probit model is fit to an item-level polychoric correlation matrix instead of a scale-level Pearson covariance matrix and estimated with weighted least squares Muthén, [4], resulting in more accurate estimates. This
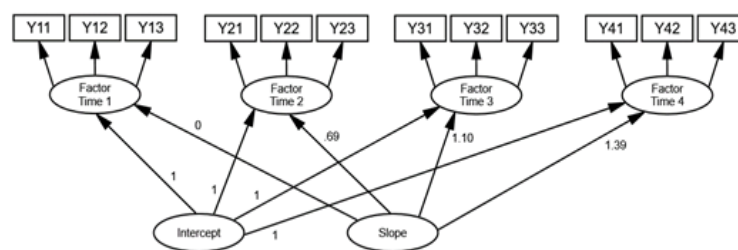
Figure 1: An illustration of latest growth curve modeling of factors.

curve-of-factors growth modeling with ordinal indicators can be efficient given appropriate sample size and sufficient computing power. Sample size requirements can also be reduced by constraining factor loadings and/or other parameters over time if measurement invariance holds.

Besides the technical modeling differences, ordinal scales may have different statistical properties, depending on the treatment and the number of choices from binary (e.g., yes/no, or true/false), trichotomous (e.g., never, sometimes, always), seven (e.g., strongly disagree, disagree, slightly disagree, neutral, slightly agree, agree, strongly agree) to even more. For instance, the mean of binary scale with 0=false/incorrect and 1=true/correct always lies between 0 and 1. In contrast, the mean of a five-choice scale, if coded as -2, -1, 0, 1, and 2, may better reflect the range of a latent variable. Although it seems plausible to sum the correct answers of dichotomous items to produce a "total correct" sum score, such sum scores have been shown to be biased against extreme cases on the latent variable dimension. Sum scores of other ordinal scales can also deviate significantly from the mean and variance of the corresponding latent variable.

Another way to treat ordinal scales in growth curve modeling is to apply weights to the different items and then average the item scores. This is intended to overcome the drawback of the implicit equal-weighting of all the items in mean/sum scores, which ignores the differential sensitivity of individual items in measuring a latent trait. There are many weighting schemes for creating composite scores [9]. For instance, maximal reliability weighting involves a confirmatory factor analysis (CFA) as the first step to identify the factor loading and residual variance of each item. The weight for each item can be generated by dividing the factor loading by the residual variance [10]. For ordinal scales, it could be sufficient to maximize the reliability of composite scores by weighting each item with its factor loading [11]. As composite scores with items weighted by factor loading are still not equivalent to the true estimates of latent variables of probability models, it remains unclear to what extent these weighted composite scores reflect the true parameters of growth curve modeling. Hereafter, we refer to this method as growth modeling of weighted means.

This study was aimed to compare the potential biases of using scale-level mean and weighted mean composite scores of ordinal items curve-of-factors in growth modeling under different numbers of response choices. We adopted the growth curve modeling of factors with ordinal indicators as the golden standard and posed no specific hypotheses about the biases of other approaches.

Method

Empirical and hypothetical population data

Two empirical ordinal datasets were used as the population data

to ensure generalizability of the findings. The first empirical dataset was extracted from an ongoing Flourishing Family Project, which was designed to monitor multifarious aspects of over 600 families of two US western areas. More information about this project can be found at (https://familycenter.byu.edu/Pages/Sponsored-Research/2007/Flourishing.aspx). Data for this study involve adolescents' ratings of their parents' psychological control on a scale of eight questions (Table 1) The participants were asked to choose one of the following options for each question: 1=never, 2=rarely, 3=sometimes, 4=often, and 5=very often.

To simulate binary and trichotomous scales, this dataset of five categories were recoded and collapsed. Specifically, never and rarely were collapsed to have a value of zero and sometimes, often, and very often were combined into a value of one. The five categories were also collapsed such that 0=never and rarely, 1=sometimes, and 2=often and very often.

The second dataset was adopted from a longitudinal project on the first-generation bilingual Chinese immigrant families with young children. These families were followed four times during a two-year period. Participants were recruited from various organizations across the Maryland-Washington DC region. An ordinal scale of maternal life satisfaction of five questions were used in this study. There are seven rating points for the participants to choose for each question, including 1=strongly disagree, 2=disagree, 3=slightly disagree, 4=neither agree nor disagree, 5=slightly agree, 6=agree, and 7=strongly agree. Thus, there were four datasets in total, three that measured psychological control respectively using binary, trichotomous, and five-point scales, and one that measured life satisfaction using a seven-point scale.

Procedure

The analysis and simulations were carried out in the following steps. First, the four empirical longitudinal datasets were subject to confirmatory factor analyses (CFA) to examine their measurement properties, including measurement invariance over time. The estimation method was weighted least square estimator with χ^2 test and degrees of freedom adjusted for the means and variances (WLSMV). The reliability (ω) for each measurement was calculated using the variance approach of McDonald [12]. We reported in detail only the CFA of the two empirical datasets of five- and seven-point scales for brevity.

Second, a latent growth curve modeling of factors (Figure 1) was respectively fit to these four datasets of binary, trichotomous, five-, and seven-point scales. The four models with their parameter estimates served as population models to generate random data for simulations.

Third, the random datasets were generated with a sample size of 300, which was presumed to yield moderate sampling variations.

This process was tantamount to drawing random samples from the population represented by the population models. A CFA was conducted with each dataset to obtain the standardized factor loadings, which were used to weight the individual items and create weighted means of the scales.

Last, latent growth curve modeling of factors (Figure 1) was fit to all these generated datasets to examine how well the population parameters can be recovered with the "golden standard" method. All new variables of the mean scores and weighted mean scores of binary, trichotomous, five-, and seven-category variables were subject to growth curve modeling for comparisons. As an exception, the weighted sums of binary scales of the simulated datasets were modeled, because the estimates of growth curve modeling were closer to the population parameters than those of weighted means.

The average estimates and standard deviations of the key parameters were compared to the original model parameters to examine potential biases. The modeling of the empirical data and simulations were conducted mainly with the latent variable modeling software Mplus (v8.0).

Result

Measurement of the ordinal scales

The measurement model of psychological control for the first data set fit the data well, with $\chi^2_{(730)}$=1909.11, p<0.01, CFI=0.94, TLI=0.93, RMSEA=0.05. The contents and factor loadings of the eight items scale are listed in Table 1. Invariance of factor loadings over time was tested by comparing this model with a model constraining the factor loadings to be equal across time. The χ^2 difference test indicated that the majority of the factor loadings were invariant over time (χ^2_{diff} (27)=38.04, p=0.08), except the last item at the first measurement that is indicated by an asterisk in Table 1 (χ^2_{diff} (1)=9.29, p<0.01). The high factor loadings and the reliabilities suggest that the psychological control was measured well over time.

The measurement model of the maternal life satisfaction in the second data set also fit the data well, with χ^2(156)=332.47, CFI=0.99, TLI=0.99, RMSEA=0.07. Factor loadings were found to be largely invariant over time (χ^2_{diff} (11)=9.45, p=0.05), except the last item at the

fourth occasion as indicated by the asterisk (χ^2_{diff} (1)=14.42, p<0.01). The item content and factor loadings are listed in Table 2. The high factor loadings and the reliabilities suggest that the construct of life satisfaction was also measured well over time. Thus, factor loadings were constrained to be invariant in subsequent latent growth curve modeling. The same tests and constraints were also applied to datasets of binary and trichotomous scales.

Latent growth curve modeling of the empirical data

A latent growth curve model with a logarithmic trajectory was identified to fit the empirical data of psychological control very well (χ^2 (730)=1909.11, p<0.01, CFI=0.94, TLI=0.93, RMSEA=0.05). As a latent construct, the initial value was set to a hypothetical mean of zero. The time scores for the model were specified as 0, 0.69, 1.10, 1.39, and 1.61, which takes the natural log of a linear trend of 1, 2, 3, 4, and 5. The growth rate was found to be α=0.13, 95% CI [0.10, 0.17], z=7.24, p<0.01, implying an upward curving change.

A similar logarithm trajectory change also fit the data of life satisfaction acceptably (χ^2 (264)=735.43, p<0.01, CFI=0.97, TLI=0.98, RMSEA=0.09). The growth rate was found to be α=-0.25, 95% CI [-0.40, -0.07], z=7.24, p<0.01, indicating a downward curving change.

Estimates of simulated data

Listed in Table 3 below are population parameters, the biases, mean estimates, and standard deviations of the five key parameters of the growth curve modeling under three different treatments of the simulated scales. A bias is defined by the difference between an average estimate of the simulated data and population parameter. The key estimates of the three treatments of the ordinal scales were compared to the population parameters with one-sample z tests.

Biases in the population parameters in Table 1 suggest the following findings. First, growth curve modeling of the factors that are measured by the ordinal scales reflected the changes of the hypothetical true population with a maximum of 0.02 differences. Biases in the variances of the intercept and slope factors and the covariance of the intercept and slope factors approximated 0.06 when using binary scales. One-sample z-tests indicated some of the population parameters can be recovered without any biases, as underlined in the table. In contrast,

Psychological Control	Age 12	Age 13	Age 14	Age 15	Age 16
Tries to change how I feel or think about things	0.51	0.46	0.48	0.55	0.63
Changes the subject whenever I have something to say	0.74	0.73	0.79	0.76	0.80
Interrupts me	0.72	0.76	0.73	0.74	0.73
Blames me for other family members' problems	0.70	0.76	0.81	0.77	0.80
Is less friendly with me if I do not see things her/his way	0.79	0.81	0.80	0.82	0.88
Brings up past mistakes when s/he criticizes me	0.72	0.73	0.76	0.78	0.79
Will avoid looking at me when I have disappointed her/him	0.77	0.74	0.77	0.74	0.86
If I have hurt her/his feelings, my parent stops talking to me until I please her/him	0.62*	0.73	0.78	0.79	0.83
Reliability ω	0.88	0.90	0.91	0.91	0.93

Table 1: Standardized factor loadings and reliabilities of psychological control scale.

	Time 1	Time 2	Time 3	Time 4
In most ways, my life is close to my ideal	0.83	0.87	0.80	0.80
The conditions of my life are excellent	0.95	0.94	0.91	0.94
I am satisfied with my life	0.99	0.96	0.90	0.96
If I could live my life over, I would change almost nothing	0.68	0.68	0.56	0.60
So far, I have gotten the important things I want in my life	0.80	0.78	0.74	0.89
Reliability ω	0.93	0.93	0.89	0.93

Table 2: Standardized factor loadings and reliabilities of the life satisfaction scale over time.

Treatment / Scales	Pop Model	Modeling Factors			Mean Scores			Weighted Mean Scores		
		Bias	M	Std.	Bias	M	Std.	Bias	M	Std.
Binary										
Intercept Mean	0.00	*0.00*	0.00	0.00	0.13	0.13	0.01	0.60	0.60	0.06
Slope Mean	0.12	-0.02	0.10	0.04	-0.09	0.03	0.01	0.02	0.14	0.07
Intercept Variance	0.27	-0.06	0.21	0.05	-0.25	0.02	0.00	0.11	0.38	0.09
Slope Variance	0.12	-0.03	0.09	0.04	-0.11	0.01	0.00	0.08	0.20	0.07
Covariance	-0.10	0.02	-0.08	0.03	0.09	-0.01	0.00	-0.03	-0.13	0.06
Trichotomous										
Intercept Mean	0.00	0.00	0.00	0.00	0.29	0.29	0.02	0.17	0.17	0.02
Slope Mean	0.05	0.01	0.06	0.01	0.02	0.07	0.02	0.00	0.05	0.02
Intercept Variance	0.24	-0.04	0.20	0.03	-0.17	0.07	0.01	-0.21	0.03	0.01
Slope Variance	0.02	*0.00*	0.02	0.00	0.02	0.04	0.01	-0.01	0.01	0.00
Covariance	-0.03	*0.00*	0.03	0.01	0.00	-0.03	0.01	0.02	-0.01	0.00
Five-Point										
Intercept Mean	0.00	*0.00*	0.00	0.00	0.84	0.84	0.04	0.47	0.47	0.03
Slope Mean	0.13	*0.00*	0.13	0.03	0.04	0.17	0.03	-0.02	0.11	0.03
Intercept Variance	0.26	-0.05	0.21	0.04	0.06	0.32	0.05	-0.15	0.11	0.03
Slope Variance	0.11	-0.03	0.08	0.02	0.04	0.15	0.04	-0.06	0.05	0.01
Covariance	-0.10	0.02	-0.08	0.02	-0.01	-0.11	0.04	0.06	-0.04	0.02
Seven-Point										
Intercept Mean	0.00	*0.00*	0.00	0.00	30.38	30.38	0.09	20.78	20.78	0.09
Slope Mean	-0.25	*0.01*	-0.24	0.05	-0.01	-0.26	0.05	0.05	-0.20	0.04
Intercept Variance	20.07	*0.03*	20.10	0.29	0.52	20.59	0.19	-0.31	10.76	0.18
Slope Variance	0.28	*0.00*	0.28	0.04	0.06	0.34	0.04	-0.04	0.24	0.03
Covariance	-0.36	*0.00*	-0.36	0.08	-0.11	-0.47	0.07	0.05	-0.31	0.06

Note: Italicized numbers indicate non-significant differences from the population parameters.

Table 3: Biases in population parameters, average, and standard deviations of growth curve modeling estimates of simulated datasets.

all the estimates of growth curve modeling of the mean scores and weighted mean scores were significantly different from the population means. Second, using the mean scores of binary scales for growth curve modeling resulted in appreciable underestimation of the slope mean, while using the mean scores of other ordinal scales for growth curve modeling reflected the change well (bias ≤ 0.04). Third, biases in the slope means of the population were similar whether using the sum of binary scales or weighted means of binary, trichotomous, or five-point scales. The appreciable bias in the slope mean occurred when using the weighted means of the seven points scale (bias=0.05). Biases in the variances and covariance of the intercept and slope factors were no better than the other two approaches. Fourth, the means of the intercept factors depended on the number of response options: the more response options, the higher the intercept means (initial levels). Fifth, average estimates of growth modeling of both means and weighted means of the ordinal scales were all significantly different from the population parameters, as the *z* tests suggested.

Discussion

This simulation study compared latent growth modeling of mean and weighted mean scores of ordinal items to full curve-of-factors modeling of the original ordinal items. The reference values for these comparisons were population model parameters derived from empirical data, so that they are more plausible and generalizable than arbitrary specifications. The change of psychological control showed a logarithmic increase, which is decelerated upward trend. This perception seemed to be reasonable as adolescents try to gain more independence and autonomy, their parents gradually increase psychological control and abandon physical and verbal coercion. As for life satisfaction of the first generation Chinese immigrants, it may be expected to decrease as adaptations to a new culture might have been accompanied by financial and job stresses.

The simulations suggested that growth modeling of factors that were measured by the ordinal scales provide good estimates of the hypothetical population parameters. Although some average estimates were significantly different from the population parameters, the magnitude of these differences are minimal, or practically trivial. In contrast, modeling the means or weighted means of ordinal items would bias the variances of the intercept and slope factors, especially the intercept factor. Large biases in variances of the intercept and slope factors could mislead practical efforts in dealing with individual differences. It is comforting that modeling the means or weighted means of ordinal scales resulted in negligible biases in the change rates of the population, except binary scales, as publications of changes estimated this way could be still credible. As means or weighted means of the ordinal scales are dependent on the number of response categories, it is difficult to compare them with the latent continuous factors measured by the ordinal scales.

Weighted means of the ordinal scales did not perform any better than the means of ordinal scales. One explanation is that weights make a difference in the composite only when the variables are not correlated. As all the scale variables are highly correlated, their contributions to the variances overlap and thus do not appear as expected [13].

This study has some limitations. We have not included other data conditions such as various distributions of the ordinal scales and sample sizes that might contribute to the biases. It was suggested by Coenders, et al. [14] that a five-point scale with middle value of zero and normal distributions could result in negligible biases in the latent variable relations, as in the case the covariance of the intercept and slope factors in this study. This is because the range of the five point scale are close to that of a typical latent variable. In addition, it could be expected that smaller samples would result in larger variances of the simulated estimates, whereas skewed distributions may result in greater

deviation from the true means. Moreover, growth curve modeling of sums of ordinal scales was not examined with simulations, because the sum differs from the mean just by a constant (division by the number of items), offering little extra information.

Another limitation of this study is that we have omitted a two-step approach (latent scoring and modeling) in the comparison. This method first obtains the estimated factor scores from measurement models and then uses these scores as observed variables in subsequent growth curve modeling [15]. This practice conforms to item response theory modeling that is widely accepted in the education field. The requirement of measurement invariance may be satisfied by testing and constraining discrimination and threshold (difficulty) parameters to be the same over time. In addition to the widely accepted theoretical basis, the advantages of this method may be less computationally time-consuming than direct modeling of the ordinal items, which may be particularly useful when a model with many items is fit to relatively small samples. This approach works well to model relations among latent constructs [16]. However, given a shortened scale of four or six items could function as well as a long one Embretson and Hershberger [17]; Kenny [18] and could be modeled directly, this approach does not appear to be advantageous for growth curve modeling, but might be examined in the future.

Conclusion

It is not advisable to use means or weighted means of ordinal items for latent growth curve modeling. Ordinal scales can best be modeled directly in latent growth curve modeling. Published reports of growth curve modeling with ordinal scales may be evaluated with findings of this study as a reference.

References

1. Stevens SS (1951) Mathematics, measurement, and psychophysics. In: Stevens SS (ed.) Handbook of Experimental Psychology, Wiley, Oxford, England, p: 26.

2. Hayes WL (1988) Statistics. Fort Worth: Holt, Rinehart and Winston.

3. Merbitz C, Morris J, Grip JC (1989) Ordinal scales and foundations of misinference. Archives of Physical Medicine and Rehabilitation 70: 308-312.

4. Muthén B (1983) Latent variable structural equation modeling with categorical data. Journal of Econometrics 22: 43-65.

5. Wright BD (1999) Fundamental measurement for psychology. In: Embretson SE, Hershberger SL (eds.) The new rules of measurement: What every psychologist and educator should know. Mahwah, NJ: Lawrence Erlbaum, pp: 65-104.

6. Reise SP, Revicki DA (2014) Handbook of item response theory modeling: Application to typical performance assessment. Taylor & Francis Group, New York.

7. Wickrama KK, Lee TK, O'Neal CW, Lorenz FO (2016) Higher-order growth curves and mixture modeling with Mplus: A practical guide. Boca Raton, FL: Taylor and Francis.

8. Muthén B, Shedden K (1999) Finite mixture modeling with mixture outcomes using the EM Algorithm. Biometrics 55: 463-469.

9. Schochet PZ (2008) Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations (NCEE 2008-4018) National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

10. Raykov T, Gabler S, Dimitrov DM (2016) Maximal Reliability and composite reliability: Examine their difference for multicomponent measuring instruments using latent variable modeling. Structural Equation Modeling: A Multidisciplinary Journal 23: 384-391.

11. Bartholomew D, Knott M, Moustaki I (2011) Latent variable models and factor analysis: A unified approach. West Sussex, John Wiley & Sons Ltd, United Kingdom.

12. McDonald RP (1999) Test theory: A unified treatment. Lawrence Erlbaum Associates. Mahwah, New Jersey.

13. Oswald FL, Putka DJ, Ock J (2015) Weight a minute… What you see in a weighted composite is probably not what you get. In: Lance CE, Vandenberg RJ (eds.) More statistical and methodological myths and urban legends.

14. Coenders G, Satorra A, Saris WE (1997) Alternative approaches to structural equation modeling of ordinal data: A Monte Carlo study. Structural Equation Modeling: A Multidisciplinary Journal 4: 261-282.

15. Hambleton RK (1991) Fundamentals of item response theory. Thousand Oaks, Sage Publications Inc, California.

16. Yang C, Nay S, Hoyle HR (2010) Three approaches to using lengthy ordinal scales in structural equation models: Parceling, latent scoring, and shortening scales. Applied Psychological Measurement 34: 122-142.

17. Embretson SE, Hershberger SL (1999) The new rules of measurement: What every psychologist and educator should know. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

18. Kenny DA (1979) Correlation and causality. Wiley, New York.