Reading Times across Sentences, Texts, and Persons: An Integrated Methodological Approach

Lee Branum-Martin and Joseph Magliano

July 12, 2020

Reading Times across Sentences, Texts, and Persons: An Integrated Methodological Approach

Lee Branum-Martin[1] & Joseph P. Magliano[2]

[1] Department of Psychology, Georgia State University

[2] Department of Learning Sciences, Georgia State University

Author Note

The authors declare that there no conflicts of interest with respect to this preprint. The authors would like to thank Eleanor Fang Yan for follow-up analyses which provided informative background to the models presented here.

Correspondence should be addressed to Joseph P. Magliano. Email: jmagliano@gsu.edu

Abstract

We present an integrated model of individual growth (multilevel SEM) to examine 10,701 reading times from 20 to 24 sentences each in four texts read by 123 college students. We evaluate the extent to which reading times indicate a single cognitive process, common across texts, versus distinct trends which suggest texts invoke different, distinctive cognitive processes. Findings suggest interesting commonalities as well as distinct aspects of sentence, text, and person-level features.

Keywords: Multilevel, SEM, Reading time

Reading Times across Sentences, Texts, and Persons: An Integrated Methodological Approach

Researchers often use sentence reading times to measure cognitive processes which yield a complex series of many responses per text and per person, best analyzed via multilevel models (Richter, 2006). A multilevel approach can estimate a person's overall speed (intercept) as well as their rate of change in speed (slope)—their tendency to speed up or slow down—as they progress through a text. The current presentation extends these ideas to multiple texts (i.e., multivariate) using multilevel structural equation models (SEMs) in order to examine the extent to which people are consistent across texts (reading times reflect a general capacity) versus people doing idiosyncratic and text-specific things for each text. Not only might people's overall speed (intercepts) be highly related across texts, but their rates of change (slopes) in progressing through different texts might also be related. We therefore present two rival models for evaluating sentence-level reading times as text-specific versus domain-general processes. In addition, we wish to evaluate the effect of person-level characteristics (reading comprehension skill) as well as text-level features. These models are presented as having general implications for examining discourse processes across multiple texts.

Method

Participants

There were 123 participants enrolled in a midwestern university. The participants received course credit for completing the study.

Materials

There were two science texts and two history texts used in the study. The two science texts had 20 sentences each and the two history texts had 23 and 24 sentences each.

Individual difference measure

Reading skill was measured with Gates MacGinitie Reading Comprehension test. Participants were given 20 minutes to complete the test.

Discourse analyses

The text sentences were analyzed for argument overlap and causal relationships. Argument overlap was determined based on whether the noun and pronouns referred back to prior nouns established in the discourse context. For each sentence, the total number of overlapping arguments was computed. To establish causal relationships, a causal network analysis was performed on each text (Trabasso, van den Broek, & Suh, 1989). Causal relationship was determined by the number of causal antecedents revealed between the sentence and prior sentences in the texts.

Procedure

Participants read the texts one sentence at a time, presented on a computer. Paragraph transitions were designated with the statement “Next Paragraph,” which preceded the first sentences of the paragraphs. Participants pressed the space bar to advance the sentences. Reading times were recorded in milliseconds, and was defined by the onset of a sentence and the pressing of the space bar to advance to the next sentences. This resulted in 87 response times per student (10,701 response times in total).

Analytic approach

Two different models were constructed to evaluate student consistency across texts, and relations to reading comprehension and text features (i.e., argument overlap and causal relations). First, we fit a parallel process model of individual growth across four texts. Intercepts and slopes for each text were allowed to freely correlate with each other. Second, we fit a single factor model of reading time across texts, which changed over time. The intercept was a single factor and the slope was a single factor, common across texts.

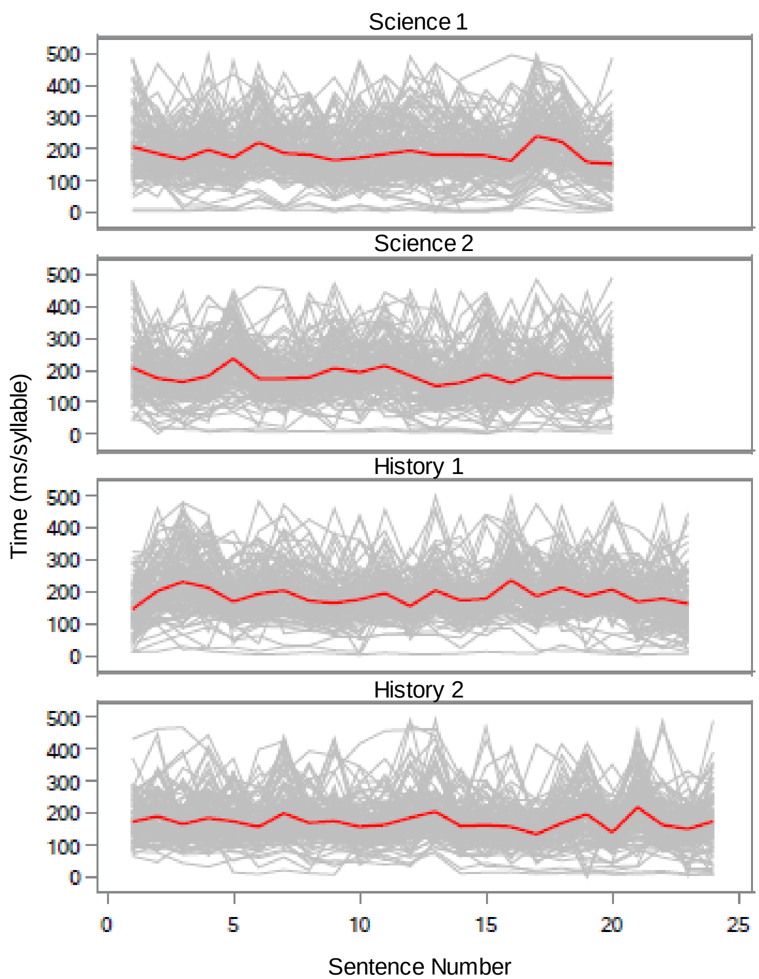
In each of these two models, reading comprehension was a person-level predictor, and causal relations and argument overlap were sentence-level predictors of reading times. Models were fit in Mplus 8.3 as multilevel SEM for time within person (TYPE = TWOLEVEL RANDOM; Muthén & Muthén, 2017).

Substantively, these two models have different foundations, based in structural hypotheses about multivariate longitudinal data (McArdle, 1988). The parallel process model is essentially four linear growth models fit simultaneously, yielding text-specific intercepts and slopes which can be examined for how they relate. The parallel process model represents the argument that reading times represent text-specific but potentially related processes.

Results

We evaluated trends across sentences for nonlinearity (Figure 1). While reading times were highly variable, there was no strong indication of consistent polynomial curvature—linear models seemed the best that could be fit for individual trends. Table 1 presents summary statistics for all variables used in the models.

Figure 1 *Individual trajectories of sentence reading times for each of the four texts*



Note. Reading times were adjusted for the number of syllables in each sentence (i.e., milliseconds per syllable). The red line in each graph is a moving average (loess regression).

Table 2 shows fit statistics for the models. The parallel process model fit better than did the single factor, both without predictors as well as in the fully conditional model, but with more than double the number of parameters (70 versus 30). The parallel process model could not be restricted to a simpler structure—correlations among slope factors were low and inconsistent (the model failed to converge). Because multilevel SEMs with random effects have no normed indices of fit, we also evaluated the results for their implications with respect to intercepts, slopes, reading comprehension, causal relatedness, and argument overlap. Therefore, results from both conditional models will be presented. Table 3 presents the correlations among the eight growth factors from the parallel process model.

Figure 2 presents the multilevel SEM diagram (Mehta, 2013) for the parallel process model (fully standardized results). The bottom section shows response times to each of the four texts (rectangles), with random intercepts and slopes for each, as a linear growth model at the person level.

Figure 3 presents the multilevel SEM diagram for the single factor growth model. The intercept and slope are formed across texts at the response level, each as a latent factor (circle) and also have a random component at the person level.

Table 1 *Zero-order correlations among all variables*

Name	Sci1	Sci2	Hist1	Hist2	S1A	S1C	S2A	S2C	H1A	H1C	H2A	H2C	Gates
Sci1	**1**												
Sci2	**0.25**	**1**											
Hist1	**0.30**	**0.23**	**1**										
Hist2	**0.23**	**0.30**	**0.26**	**1**									
Sci 1 Arg	**-0.14**	-0.02	0.11	0.01	1								
Sci 1 Cause	**0.05**	0.06	0.00	0.05	-0.13	1							
Sci 2 Arg	-0.06	**-0.05**	0.11	0.14	0.1	0.12	1						
Sci 2 Cause	-0.04	**-0.02**	-0.06	-0.06	-0.11	-0.08	0.05	1					
Hist 1 Arg	-0.04	-0.08	**0.01**	-0.07	0.08	-0.19	-0.23	-0.05	1				
Hist 1 Cause	-0.05	0.00	**-0.10**	-0.03	0.03	-0.12	-0.26	0.33	-0.01	1			
Hist 2 Arg	0.06	-0.05	0.12	**-0.04**	0.21	0.27	-0.14	-0.04	-0.06	-0.11	1		
Hist 2 Cause	0.00	-0.05	0.01	**-0.09**	0.05	-0.34	-0.16	0.43	0.21	0.41	-0.06	1	
Gates	**-0.09**	**-0.1**	**-0.13**	**-0.1**	0	0	0	0	0	0	0	0	1
n	2416	2437	2801	2930	2416	2416	2437	2437	2801	2801	2930	2930	2952
mean	185.6	183.4	187.2	171.9	2.6	0.9	2.6	1	2	1.3	2.5	0.8	27.4
SD	84.5	80.8	76.8	72.5	1.6	0.7	1.2	0.9	1.1	1.1	1.6	0.8	7.8
min	1.6	1.1	2.9	1.6	0	0	1	0	0	0	0	0	10
max	495.7	490.4	499.9	493.5	7	2	6	3	4	3	6	2	47

Note. n = number of reading times across sentences (participants = 123). Results are presented for two science (Sci) and two history (Hist) texts. Each sentence was evaluated for argument overlap (Arg) and causal connections (Cause). Gates = Gates MacGinitie passage comprehension score.

Table 2 *Model Fit*

	Model	Parameters	Deviance
1	Parallel Process	54	117,168
2	One Factor	20	117,327
1a	Parallel Process, Conditional	70	103,640
2a	One Factor, Conditional	30	103,862

Table 3 *Correlations among person growth factors, Parallel process model.*

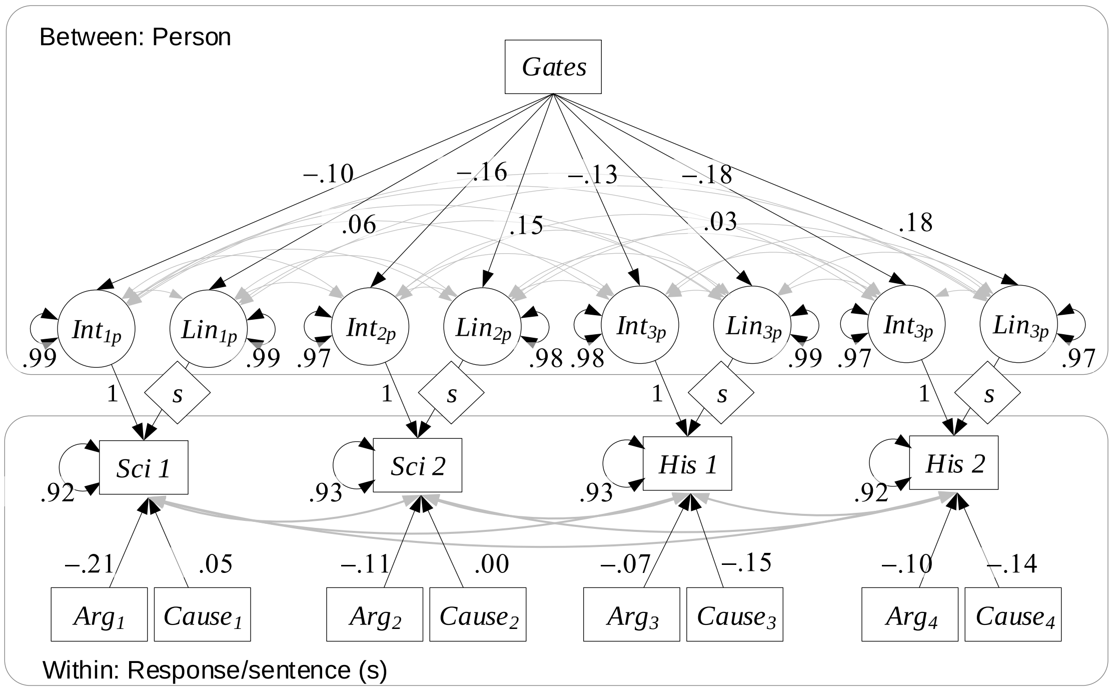
	Sci1	Sci2	Hist1	Hist2	Lin1	Lin2	Lin3	Lin4
Sci1	1							
Sci2	0.58	1						
Hist1	0.71	0.56	1					
Hist2	0.38	0.55	0.53	1				
Lin1	-0.28	-0.18	-0.25	0.07	1			
Lin2	-0.07	-0.52	0.01	-0.11	0.03	1		
Lin3	-0.05	-0.05	-0.49	0.01	0.38	0.04	1	
Lin4	0.17	-0.03	0.15	-0.48	-0.14	0.48	-0.11	1
Mean	191	193	191	179	-0.36	-0.86	-0.26	-0.51
Bet. SD	65	64	59	45	2.17	2.82	1.87	2.15
W/in SD	61	61	60	57	—	—	—	—

Intercepts: are slow readers generally slow? The parallel process model showed intercepts were moderately related, but could not be forced to be unitary. The single factor model had strong loadings, but lower fit overall. Both models had similar overall predictions.

Slopes: how consistently do readers speed up or slow down? The parallel process model showed changes in rates were highly inconsistent across texts. The single factor model slope factor was weak, with low standardized loadings, and low variance (and poorer fit overall). There was little overall consistency in rates of change across sentences in different texts.

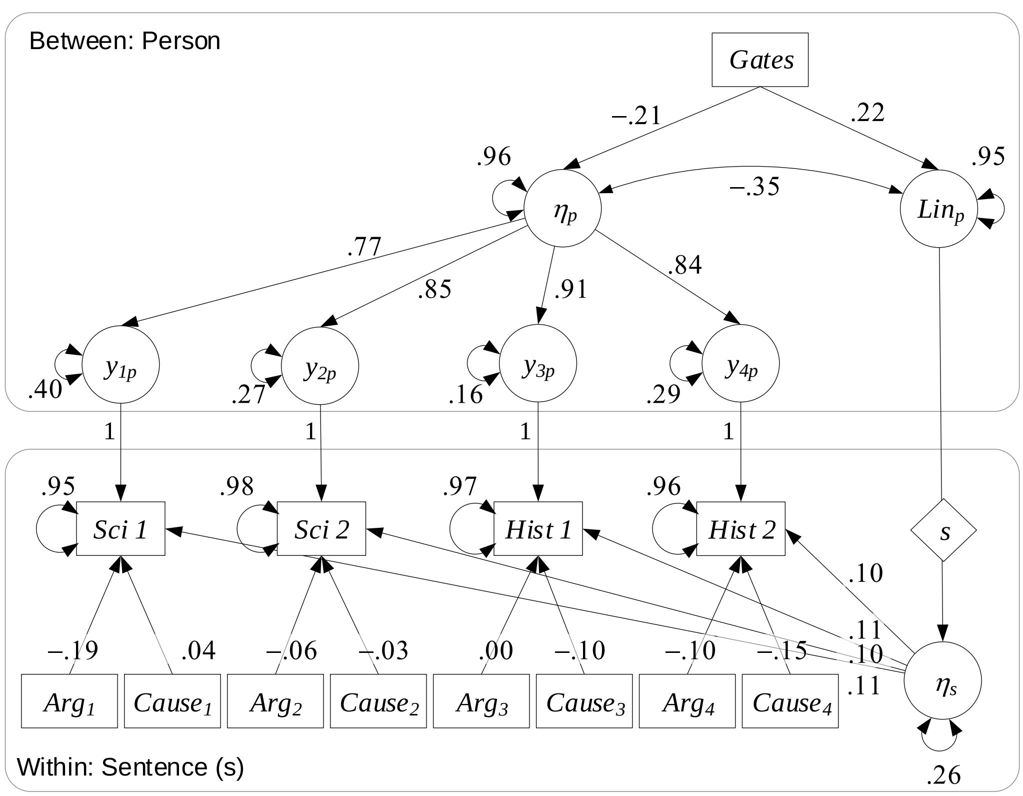
Text features. Argument overlap sped up reading times in both models (β = .00 to -.21). Causal relations sped reading in history (β = -.10 to -.15) but not in science (β = -.03 to .04). These features had similar estimated effects in both models.

Figure 2 *Parallel Process Multilevel SEM*



Note. Person-level cross-text correlations shown in Table 3.

Figure 3 *Single Factor Multilevel SEM*



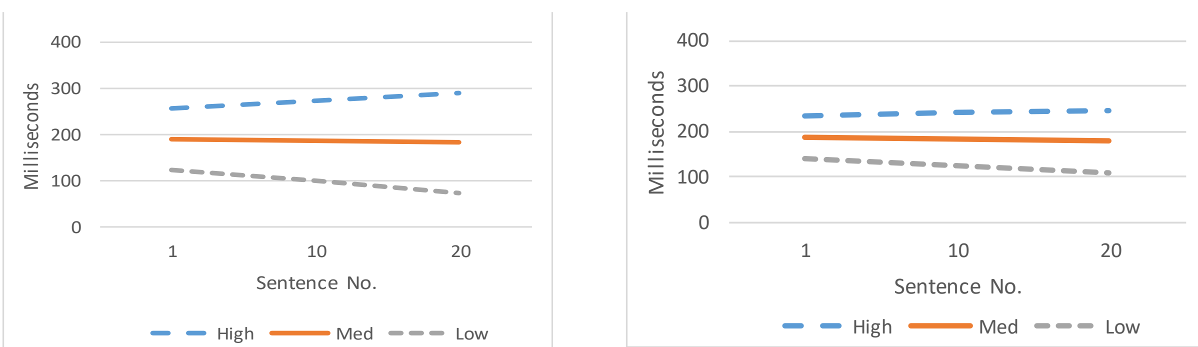
Model-predicted trends. Figure 4 shows the model-predicted trajectories of reading times for the parallel process model (A) and the single factor model (B). The solid line indicates the model-predicted mean, with a trajectory for high slope (+1 SD) slowing down (large dash) and a low slope (-1 SD) speeding up through the text. Panel B shows the same model-predicted trajectories for the single factor model.

Person level reading comprehension. The parallel process model showed better readers were faster (r = -.10 to -.18), and there was little relation of slopes across texts (r = .03 to .18). The single factor model was similar, better readers were faster (r = -.21) and tended to slow down (r = .22) through the text. Model-predicted trajectories were highly similar across the two models and across the four texts.

Figure 4

Model-predicted Trajectories

A *Parallel Process Model* **B** *Single Factor Model*



Note. Science text 1 shown as an example.

Discussion

The results of this study indicate that relative reading rates of participants were consistent across texts. As such, slow or fast readers tend to read similarly in each text. In the parallel process model, intercepts were correlated .38 to .71. In the single factor model, standardized loadings (validity coefficients) were .77 to .91. Moreover, reading comprehension was negatively correlated with sentence reading time intercepts (-.10 to -.21 in each model), suggesting that better readers were faster.

However, we found that the slopes were not consistently related, and consistency of intercepts was not sufficient to override the lack of relation among slopes (i.e., a simpler model with fewer factors was not appropriate). This suggests that changes in reading rates are not consistent across texts and appear idiosyncratic to that combination of text and reader. For example, a reader may slow down as they read one text, but for another text they may speed up as they progress through sentences.

In terms of text features, we found that argument overlap was usually related to decreases in sentence reading times, whereas causality was only significantly predictive of a decrease in sentence reading times for history (not for science).

The results of this study suggest that assumptions that the relative rate of reading is consistent across texts in other studies appears generally correct (intercepts are reasonably related). In addition, better readers are generally faster. However, these consistency effects are not perfect. Moreover, there may be text-specific factors that affect how reading rates change across sentences, and univariate or single-outcome approaches such as multiple regression (single or multilevel) may not be well suited to model these text specific effects or possible relations across multiple texts.

It must be noted that there is some ambiguity about using growth or multilevel modeling for sentence reading times. The metric of sentences is not necessarily consistent, as is required for a growth model over a set metric of time (e.g., weeks or years). To the extent that not every sentence is equal, examining "growth" across numbers of sentences might not have a consistent, stable metric in the same way that students' reading speed (outcome) every week (time) would. In addition to this conceptual and statistical ambiguity, the individual student time trajectories are not well captured. Standard linear and polynomial growth models do not appear to fit well (in standard individual growth SEMs). It is possible that sentence reading times are not systematic enough to be adequately captured by complex multilevel linear models such as those used here. Individual level growth models as well as these multivariate growth models will have to be tested in larger samples with more texts, but these preliminary findings suggest reading times might not be as systematic or cohesive as needed for multilevel linear SEM and its simpler versions (e.g., standard growth models).

The model-based trajectories and estimates for text features and comprehension relations are highly similar in both models. To the extent that there is commonality across texts and rates, the single factor model is highly parsimonious, possibly at the expense of being overly general. For modeling, these results suggest that stacking all four passages into a single, omnibus multilevel model will miss many differences across texts, but the overall text and person level relations may remain reasonably consistent with those reported here. Overall, the features of consistency versus difference will need to be tested beyond the four texts and limited sample used here.

References

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561-614). New York: Plenum Press.

Mehta, P. D. (2013). n-level structural equation modeling. In Y. Petscher, C. Schatschneider, & D. L. Compton (Eds.), *Applied quantitative analysis in the social sciences* (pp. 329-362). New York: Routledge.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide. Eighth edition.* Los Angeles, CA: Muthén & Muthén.

Richter, T. (2006). What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes: A Multidisciplinary Journal*, *41*(3), 221-250.

Trabasso, T., van den Broek, P., & Suh, S. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, *12*, 1-25. doi:10.1080/ 01638538909544717