Applying the Measurement Model of Derivatives to Evaluate and Refine Measurement Scales in Longitudinal Education Data Using the ECLS-K

Kyle D. Husmann, Timothy R. Brick, and James C. DiPerna
Pennsylvania State University

Abstract

The Measurement Model of Derivatives (MMOD; Estabrook, 2015) provides the opportunity to evaluate and refine measurement scales used in longitudinal education data to clarify their theoretical distinctions and relationship to academic achievement. We demonstrate this using three teacher-rated scales of child self-regulatory behavior obtained from the Early Childhood Longitudinal Study Kindergarten Class of 2010-11 (ECLS-K:2011; Tourangeau et al., 2019). Theoretical measurement structures are compared to data-driven structures on a holdout sample, then compared in their ability to predict academic achievement on a validation sample to externally validate their utility. We discuss theoretical implications for self-regulation, as well as the MMOD's applicability to other educational data sets.

Keywords: measurement model, longitudinal data analysis, ECLS-K:2011, educational psychology

Kyle D. Husmann (corresponding author) is a pre-doctoral student in Human Development and Family Studies at Pennsylvania State University, 115 Health and Human Development, University Park, PA 16802, USA; e-mail: kdh38@psu.edu. His research focuses on novel methods that can help understand the processes and mechanisms that shape children's academic trajectories across multiple timescales.

Timothy R. Brick is an assistant professor of Human Development and Family Studies at Pennsylvania State University, 115 Health and Human Development, University Park, PA 16802, USA; e-mail: tbrick@psu.edu. His research interests focus on the development of methods for modeling and analyzing intensive longitudinal data and studying the dynamics of intra- and interpersonal systems as they unfold in real time.

James C. DiPerna is Professor in Charge of School Psychology at Pennsylvania State University, 103 CEDAR Building, University Park, PA 16802; jdiperna@psu.edu; e-mail: jdiperna@psu.edu. His research interests focus on the development of school-based assessments and interventions to enhance students' social, emotional, and academic competence.

The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305B090007 to The Pennsylvania State University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Introduction

Over the past quarter-century, longitudinal investigations have become more prevalent in the field of educational research, partially to address salient questions in education relating to the processes by which skills and behaviors develop and change over time. Since these processes often have multilevel structure poorly described by cross-sectional analyses (Molenaar, 2004), longitudinal designs have become an essential pillar of modern educational research.

This shifting focus has led to the development of a variety of large-scale longitudinal studies for the education research community. Yet despite the increasing use of longitudinal analyses, the measurement scales selected and developed to collect data for these studies are often evaluated only cross-sectionally, using techniques like standard confirmatory factor analysis (CFA; Jöreskog, 1969). Extensions of CFA, such as longitudinal factor analysis (Widaman et al., 2010), can assess the longitudinal factor invariance of latent constructs across time but the trajectories of the manifest variables across time are not taken into account. As a consequence, items appearing to belong to the same construct at multiple points in time may still be exhibiting qualitatively different trajectories across time, indicating that the given measurement structure was not a good match for the underlying developmental processes.

Measurement models that accurately reflect underlying developmental processes are important for understanding how skills and behaviors change over time. In educational research, accurate models of developmental processes can help clarify predictors of academic achievement and illuminate points of leverage that interventions can target to support individuals' academic progress. Existing work assessing which cognitive and socioemotional indicators are predictive of later academic performance (Claessens et al., 2009; DiPerna et al., 2007; Duncan et al., 2007; Li-Grining et al., 2010; Pan & Zhu, 2018; Sung & Wickrama, 2018) relies on the assumption that the measurement structures of the scales reflect the longitudinal dynamics of the underlying processes driving academic performance. The extent to which this assumption does not hold will attenuate relationships between these indicators and later outcomes. Furthermore, because these studies often use scales to assess related constructs from an assortment of measures, any content overlap of items across scales will additionally attenuate their ability to differentially predict outcomes of interest. The closer measurement structures can reflect underlying developmental processes, the more useful they will be to inform interventions and policy intended to shape developmental trajectories towards more positive outcomes.

The Measurement Model of Derivatives (MMOD; Estabrook, 2015) is a recently developed method for assessing the utility of measurement models of developmental constructs on multivariate data with a small number of repeated measures. The MMOD assesses the measurement structure of items as well as their temporal dynamics by leveraging the rates of change (mathematically, the derivatives) of item levels across time. In this paper we show how the MMOD can be used to evaluate and refine the measurement structures used to study longitudinal processes in education so that they accurately reflect the developmental nature of the processes they represent.

Background

The Early Childhood Longitudinal Study Kindergarten Class of 2010-11 (ECLS-K:2011; Tourangeau et al., 2019), is a widely-used educational data set with a nationally representative sample. The ECLS-K:2011 followed children from kindergarten through fifth grade surveying their early academic experiences from the perspectives of parents, teachers, and individual assessments. While the data collected in the ECLS-K:2011 are longitudinal, the measurement instruments selected and developed for the study have only been evaluated using traditional psychometric methods that do not consider item-level temporal dynamics (Najarian et al., 2017, 2018).

Self-Regulation and the ECLS-K:2011

Teacher-reported measures of child self-regulatory behavior with publicly accessible item-level data from the ECLS-K:2011 provide an opportunity to demonstrate the utility of the MMOD in evaluating and refining measurement structures in longitudinal education data. Self-regulation, defined broadly, is the ability to direct or control one's thoughts, feelings and actions (McClelland & Cameron, 2012). While self-regulation is widely recognized as an essential part of a child's academic development, a critical aspect of school readiness, and a predictor of academic achievement (Blair & Raver, 2015), there is less agreement on how to define and measure the construct and its underlying components (Blair & Razza, 2007; Duncan et al., 2007; McClelland & Cameron, 2011; McClelland et al., 2007).

In constructing the ECLS-K:2011, researchers selected scales from several sources to measure different aspects of self-regulation. For the present study, we use only teacher-reported measures of child self-regulatory behavior with publicly available item-level data. By focusing the investigation on teacher report we avoid the additional complexity in interpretation when combining test items from multiple contexts and perspectives.

Evaluating and refining these measurement structures can clarify theoretical distinctions between constructs of self-regulation within and across research fields and perspectives. For example, if we find fewer factors than we expect when pooling items across measurement scales, it would suggest that we are making a distinction between two constructs that are actually redundant. Similarly, if we find more factors than we expect, it would suggest that we are conflating two separable constructs. If our chosen factor structures are better predictors of related academic outcomes on an independent holdout data set, this provides convergent evidence that our factor structures are better matches for the underlying processes at work. Theory that most accurately represents underlying developmental processes and their relationship to academic achievement is important for designing successful educational interventions. Clearer theoretical distinctions between related constructs in the study of self-regulation can help to build and inform school readiness programs that include curricula to specifically target self-regulatory skills, like the Head Start REDI program (Bierman et al., 2008).

Structural Validity

Structural validity, a critical component of construct validity, is a reflection of how well a test's item structure maps to the structure of the latent traits and is related to the unidimensionality or homogeneity of each measurement scale (Loevinger, 1957). Cronbach's

alpha coefficient (Cronbach, 1951), a measure of internal consistency, is often reported in this context, but this alone is not sufficient for demonstrating unidimensionality (Clark & Watson, 1995; Cortina, 1993). Confirmatory Factor Analysis (CFA; Jöreskog, 1969) is most commonly used to evaluate the structure of a measurement scale, but as standard CFA operates cross-sectionally, there can be ambiguity in differentiating between factor structures.

Figure 1 illustrates an example provided by Estabrook (2015) of two confirmatory models that are statistically equivalent when evaluated at a single time point, but have very different substantive interpretations. The model in Figure 1a hypothesizes a single latent construct that explains the variance in the manifest variables, with a residual covariance between two of the items, while the model shown in Figure 1b hypothesizes two latent oblique factors. Because these models make identical predictions, the final measurement model must be chosen using a priori theory and cannot be supported or rejected by real-world data. The extent to which the structure of the actual underlying processes does not match the structure of the chosen measurement model, the model's utility in predicting or suggesting points of leverage to manipulate outcomes of interest will be degraded.

Figure 1

Two models with structural differences that make identical predictions

To statistically compare the validity of these two measurement structures, additional information is needed. By extending these models to include an additional measurement occasion (Figure 2), temporal variation at the item-level is leveraged to statistically differentiate between the two substantively different model structures. In this example it is straightforward to extend the cross-sectional model into a longitudinal context to model item-level temporal variation. As measurement occasions are added and measurement models become more complex, however, the solution is not always as clear.

Dynamic factor analysis (DFA; Molenaar et al., 1992) is an extension of factor analysis that explicitly takes temporal dynamics into account when evaluating item structure. However, DFA requires a large number of measurements per participant, which is why it is typically used for time series data with few participants and few measures. Recent adaptations (Tripodis & Ziropiannis, 2015) have enabled the technique to be used in panel studies with many participants, many measures, and few measurement occasions. Still, as

(a) A one factor model with a residual covariance

(b) An oblique two factor model with simple structure

Figure 2

Model structures can be statistically differentiated by adding an additional measurement occasion

DFA only includes timepoint-to-timepoint item autocorrelations in the factor structure, it cannot account for the full complexity of temporal dynamics that may be present in the data.

The Measurement Model of Derivatives (MMOD; Estabrook, 2015), by contrast, explicitly models item trajectories by creating latent variables that directly represent mathematical derivatives of a construct. These derivatives directly quantify the change in that construct across time. By including a representation of change in the model, latent derivatives allow researchers to precisely articulate theories about developmental change (Deboeck et al., 2015).

The first three derivatives of a construct with respect to time have intuitive substantive interpretations: level, velocity and acceleration. These quantities are easiest to intuitively understand in the context of driving a car. The zeroth derivative, or level, is simply the value of the construct at a moment in time; the mile marker the car is passing. The first derivative, or velocity, is the rate that the construct's level is changing at that

moment in time; the speed of the car as indicated by its speedometer. The second derivative, or acceleration, is the rate that the construct's velocity is changing at that moment in time, which is manipulated in the car by pressing or releasing the gas pedal. Higher order derivatives follow the same pattern in which each represents the rate of change of the previous derivative. It is rare for derivatives higher than acceleration to be of particular interest in behavioral studies.

Figure 3

A one-factor MMOD with four manifest variables at three measurement occasions

Figure 3 shows the construction of a one-factor MMOD built with four manifest variables at three measurement occasions. The MMOD estimates latent derivatives of manifest variables using paths weighted by Generalized Orthogonal Local Derivative estimates (GOLD; Deboeck, 2010). Substantively, this captures the average level, slope, acceleration and higher order change dynamics of items across their measurement occasions. In this way, the weights of the derivative transform can also be compared to coefficients in a growth modeling framework, where the zeroth derivatives capture differences in mean level, the first derivatives capture linear change, the second derivatives capture quadratic change and higher order derivatives capture higher order polynomial change of manifest variables across measurement occasions.

The MMOD works well with panel data of 3-5 measurement occasions and can easily be built with existing, free, off-the-shelf SEM analysis software, like OpenMx (Neale et al., 2016) or lavaan (Rosseel, 2012).

Present Study

In the present study, we demonstrate the utility of the MMOD for evaluating the structure of items on a multi-faceted educational assessment along with their temporal dynamics. We first generate a pool of candidate factor structures using exploratory factor analyses on a training sample, then use the MMOD on a test sample to select the best-

fitting candidate item structure. The selected best-fitting candidate item structure is also compared to the theoretical item structure using the MMOD on a test sample. Finally, the selected structure and the theoretical item structure are compared in their ability to predict academic achievement on a separate validation sample.

Method

Approach

The goal of this study is to determine the extent to which the theoretical item structures of teacher reported child self-regulation from the ECLS-K:2011 match the structure and dynamics of the collected longitudinal data. Instead of employing a traditional cross-sectional factor analysis, the MMOD is used in order to take item-level temporal dynamics into account. The evaluation is performed in three parts: 1) derive candidate data-driven factor structures at each measurement occasion using a training subset of the complete ECLS-K:2011 data set, 2) choose the best performing candidate measurement structure according to MMOD fit and compare its fit to an MMOD of the theoretical factor structure using an independent ECLS-K:2011 subset (which we call the *test* subset) 3) externally verify the utility of the MMOD's evaluation by comparing the ability of the MMOD-selected factor structure to the theoretical structure in its ability to predict academic achievement using another subset of the ECLS-K:2011 independent of the other two (the *validation* subset). The three data subsets—training, test, and validation—are independent subsets drawn from the complete data set before analysis, so that each step can be performed on an independent sub-sample of the ECLS-K:2011. This use of hold-out testing and validation data sets reduces the chances that the findings are the results of overfitting to a single data set, and thereby improve the generalizability of the model (Arlot & Celisse, 2010).

Sample

As noted previously, the ECLS-K:2011 is a nationally representative sample of children followed from kindergarten through the 5th grade. It includes interviews with parents and teachers as well as direct assessments of the children to capture their early educational experiences through a diverse range of perspectives and methods. The ECLS-K:2011 is a natural choice to demonstrate the use of the MMOD on longitudinal education data because it has publicly available item-level data on multiple measurement scales over multiple occasions.

Children were excluded from the data set for this study if they were not first-time kindergarteners or in the same school at the three measurement occasions used in the analyses. The total sample size, after applying these selection criteria, was 11286.

Cases from the study data set were then randomly divided into three parts in a 25-25-50 split: a training subset (N=2821) to run exploratory analyses to generate data-driven factor structures, a test subset (N=2822) used to compare the theoretical and data-driven measurement structures in a confirmatory MMOD, and a validation subset (N=5643) set used to compare the ability of the measurement structure to predict academic achievement. No single row of data appears in more than one subset in order to ensure the independence of the data subsets.

Measures

Three teacher-reported measurement scales collected by the ECLS-K:2011 related to self-regulation were chosen for this study: Approaches to Learning (ATL), Attentional Focusing (AF), and Inhibitory Control (IC). These scales were chosen because they were available from the first measurement occasion (fall of kindergarten) and had item-level data publicly available. Standardized math and reading scores were used as outcome measures.

Three measurement occasions included all of these scales using the same survey items from the beginning of the ECLS-K:2011: fall of kindergarten, spring of kindergarten, and spring of first grade. After the spring of first grade measurement occasion, the ECLS-K:2011 changed the AF and IC measurements to insure items remained age-appropriate. Because we are studying item-level dynamics across time and require responses from identical items across measurement occasions, the present study uses only these three measurement occasions.

Approaches to Learning

As measured by the ECLS-K, Approaches to Learning (ATL) reflects domain-general behaviors related to a child's ability to self-regulate their learning process and environment. The scale was originally developed for the ECLS-K 1998-99 cohort (Tourangeau et al., 2009) and administered at every measurement occasion. Items were rated on the frequency of the observed behavior on a scale from 1 (never) to 4 (very often). The scale also included an option to indicate if the behavior was not observed which was converted to a missing value for the present study. The ATL scale demonstrated high reliability, with a Cronbach's alpha coefficient of 0.91 for all three measurement occasions selected for this study.

Attentional Focusing and Inhibitory Control

Attentional Focusing (AF) and Inhibitory Control (IC) were each measured using six items from the Short Form of the Children's Behavior Questionnaire (CBQ; Putnam and Rothbart, 2006). Rothbart et al. (2001) defined AF as a child's ability to stay on task and complete activities, and IC as their ability to resist impulsive behaviors. Items describing behaviors like "Does the child have trouble sitting still when told?" were rated on a 7-point scale from "extremely untrue" to "extremely true". Teachers could also respond with "not applicable" if they did not feel the item applied to the child, which was converted to a missing value for the present study. Both the AF and IC scales had high internal reliability for each measurement occasion selected for this study, with Cronbach's alpha coefficients ranging from 0.83 to 0.87.

Reading and Mathematics IRT Scores

Reading and Mathematics achievement tests were administered on the direct cognitive assessment portion of the ECLS-K:2011. The ECLS-K:2011 used Item Response Theory (IRT) to calculate standardized scores for children such that each child's ability could be estimated on a common scale and evaluated longitudinally, even though children received different test items at each measurement occasion. For ease of interpretability, the present study uses the theta scores provided by the ECLS-K:2011 which were computed using the entire data set after the final round of data collection. Theta scores are overall

scales of ability that can be used to make comparisons between and within children across measurement occasions. The complete item pool for the reading and math assessments contained 205 and 206 items respectively. Both the reading and math assessments had similarly high internal reliability over the measurement occasions used in this study, with Cronbach's alpha coefficients ranging from 0.92-0.95.

Analysis

R version 3.5.1 (2018-07-02) (R Core Team, 2019) and OpenMx 2.16.0 (Neale et al., 2016) were used for all analyses. MMODs were built using mxmmmod 1.0.1 (Husmann et al., 2020). Snakemake 5.5.4 (Köster & Rahmann, 2012) was used to manage the analysis workflow and environment. All of the results in this study can be reproduced using the code found in (Husmann, 2020).

Candidate Factor Structures

Exploratory factor analyses were conducted with the training data set to generate data-driven candidate factor structures. Any missing data were line deleted from this part of the analysis. Manifest variables that received over 20% of "not applicable" or "not observed" responses at any occasion were removed from consideration in the candidate structures and subsequent analyses. At each time point, exploratory analyses of 2, 3, 4, and 5 factors were run to create candidate data-driven factor structures to be tested with the MMOD. Factor solutions were rotated to simple structure using *promax* rotation (Hendrickson & White, 1964), an oblique factor rotation which permits both orthogonal and non-orthogonal factor structures. Items were assigned to a factor parcel if that factor had the loading with the largest absolute value.

MMOD Model Evaluation

Every candidate factor structure generated by the training data set was validated by an MMOD on the testing data set. Missing data were accommodated using full-information maximum likelihood. Model fit parameters were calculated, and the Bayesian Information Criterion (BIC) was used to choose the best fitting factor structure.

Predicting Academic Achievement

The best-fitting factor structure was then compared to the theoretical structure in its ability to predict academic achievement. For each candidate solution, factor parcel scores were computed for each factor by averaging each child's scores on the items assigned to that factor. This approach matches the scoring procedure used by the ECLS-K:2011 for the theoretical ATL, AF, and IC scales (Tourangeau et al., 2019). Factor parcel scores were grand mean centered and scaled to unity variance to aid in the interpretation of results. Missingness was handled by case-wise deletion (including incomplete trajectories did not change the pattern of results). The resulting scores were used to predict both reading and math scores in the following multilevel model:

$$Y_{ti} = \pi_{0i} + \pi_1 T_{ti} + \pi_{F_k} F_{kti} + e_{ti} \quad (1)$$

$$\pi_{0i} = \beta_{00} + u_{0i} \quad (2)$$

Where Y_{ti} is the math or reading IRT for individual i score at occasion t ; β_{00} and π_1 are the fixed effects intercept and slope; u_{0i} is the random effects intercept; π_{F_k} are the fixed effect of each time-varying factor on achievement; and e_{ti} is measurement error.

Results

Candidate Factor Structures

Five unique candidate factor structures were obtained using exploratory factor analysis at each measurement occasion on the training data set: one 2-factor structure, two 3-factor structures, and two 4-factor structures. Structures with 5-factors were not considered as candidate factor structures for the MMOD because each contained at least one factor with fewer than three items, which are too few to uniquely identify a factor.

A full list of factor loadings and correlations for the candidate structures can be found in the Appendix. Factor structures showed a few consistent patterns. In general, strong factors were commonly identified corresponding to the three hypothesized constructs (ATL, AF, and IC) and labeled accordingly. Additionally, a common factor emerged comprising three items from the AF scale. This 3-item factor, which we labeled as "Behavioral Engagement" (BE; see Discussion, below), was found in all of the 2-, 3-, and 4-factor structures. The 2-factor structure comprised of the 3-item BE factor along with a catch-all factor with the remaining items. In the 3-factor structures, this catch-all factor split into an ATL factor along with a AF/IC factor. In the 4-factor structures the AF/IC factor differentiated into separate AF and IC factors.

In all structures, the BE factor was less correlated to the other factors. In the 4-factor structures, for example, ATL, AF and IC have correlations ranging from 0.71 to 0.79, while correlations with the BE factor range from 0.47 to 0.67. For a full list of the factor correlations for the candidate factor structures, see the Appendix.

Measurement Model of Derivatives

The MMOD was fitted to each candidate factor structure. Table 1 shows the fit statistics for each MMOD fit on the test data set. Every data-driven factor structure had a better fit on the test data set across all fit metrics than the theoretical structure as defined by the original measurement instruments. Of the five candidate factor structures, the 4-factor structure found at the fall and spring of kindergarten provided the best MMOD fit, which we subsequently refer to as the MMOD-selected factor structure.

Predicting Academic Achievement

To demonstrate the utility of the MMOD-selected factor structure, we computed factor parcel scores by averaging the items attributed to each factor. These scores were then used as predictors in the multilevel regression predicting reading and math achievement (as indicated by IRT theta scores). Table 2 compares the multilevel models predicting achievement from parcel scores derived from the theoretical 3-factor and MMOD-selected factor structures. The magnitude of the significant regression coefficients for the factors

Table 1
MMOD Model Fit Statistics of Candidate Factor Structures on Test Subset

Structure Type	Structure Source	AIC	BIC	rmsea	cfi	tli
2-Factor	Fall Kindergarten					
	Spring Kindergarten	79145	-697394	0.051	0.912	0.907
	Spring 1st Grade					
3-Factor	Fall Kindergarten	77633	-698781	0.047	0.927	0.922
	Spring Kindergarten					
	Spring 1st Grade	77999	-698415	0.048	0.924	0.918
4-Factor	Fall Kindergarten					
	Spring Kindergarten	76594	-699641	0.044	0.938	0.931
	Spring 1st Grade	76988	-699247	0.045	0.934	0.927
3-Factor	Theory	80119	-696294	0.054	0.903	0.896

Note: According to the MMOD, the best-fitting factor structure was the 4-factor structure found in both the Fall and Spring of Kindergarten measurement occasions (fit statistics in bold)

ranged from 0.02 to 0.15. These scores are all meaningful in practice considering the outcome theta scores are in standard deviation units, standardized across grades K-5. Even the smallest (.02) means that a student one SD above the mean steps ahead of about 80 additional students in the sample compared to an identical student at the mean. At the higher levels (e.g. .11), one standard deviation of difference in the predictor can be loosely equated to one month of schooling.

The MMOD-selected 4-factor structure was superior across all fit measures in predicting reading scores as compared to the theoretical structure (BIC = 15571 and BIC = 15595 respectively). For predicting math scores, the MMOD-selected and theoretical structures provided similar fit (BIC = 11586 and BIC = 11588). Importantly, BE was a significant predictor of both reading and math achievement, indicating that it contributes unique information over and above the contributions of the other three factor parcels.

Discussion

Using the items from three teacher-reported measures of self-regulation from the ECLS-K:2011, we generated five candidate factor structures using traditional cross-sectional exploratory factor analysis on a training subset of the ECLS-K:2011. These structures were then tested longitudinally using the Measurement Model of Derivatives (MMOD) on a separate testing subset of the ECLS-K:2011. Because the MMOD fit includes item-level temporal dynamics, it provides information about how well a given factor structure matches the underlying developmental processes.

The candidate factor structure with the best MMOD BIC fit, or the "MMOD-selected factor structure", was a 4-factor structure. The MMOD fit of this 4-factor structure

Table 2

Results from multilevel models predicting reading and math achievement on the validation data set using the 4-factor structure chosen by best MMOD fit as compared to the theoretical factor structure

	Reading Theta Score				Math Theta Score			
	Theory 3 Factor		MMOD 4 Factor		Theory 3 Factor		MMOD 4 Factor	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Factor Fixed Effects								
Approaches to Learning	0.135*	0.010	0.150*	0.009	0.081*	0.007	0.088*	0.007
Attentional Focusing	0.093*	0.009	0.050*	0.009	0.056*	0.007	0.031*	0.007
Inhibitory Control	-0.043*	0.009	-0.048*	0.009	-0.013	0.007	-0.018*	0.007
Behavioral Engagement			0.045*	0.006			0.030*	0.005
Other Fixed Effects								
(Intercept)	-0.981*	0.009	-0.979*	0.009	-0.947*	0.009	-0.946*	0.009
Grade	1.114*	0.006	1.111*	0.007	1.042*	0.005	1.040*	0.005
Random Effects								
sd(Intercept)	0.422		0.419		0.466		0.465	
sd(resid)	0.407		0.407		0.299		0.299	
Model Fit								
LL	-7765		-7748		-5762		-5756	
df	10202		10201		10182		10181	
AIC	15544		15513		11538		11528	
BIC	15595		15571		11588		11586	

also outperformed the MMOD fit of the theoretical structure by all statistical fit measures including AIC, BIC, RMSEA, and CFI/TLI, suggesting that the MMOD-selected structure was a better match for the underlying developmental processes. The first three factors in the MMOD-selected factor structure generally aligned with the theoretical factor structure, regrouping items into Approaches to Learning (ATL), Attentional Focusing (AF), and Inhibitory Control (IC) factors with only a few items migrating away from their theoretical factor. The fourth factor consisted of three items originally from the theoretical structure's Attentional Focusing (AF) scale:

- "Sometimes becomes absorbed in a picture book and looks at it for a long time"
- "When building or putting something together, becomes very involved in what s/he is doing, and works for long periods", and
- "When drawing or coloring in a book, shows strong concentration"

The emergence of this new factor suggests that the theoretical AF scale may be conflating two distinct developmental constructs: AF, and something else measured by the three items above. These three items all measure the extent to which students are participating in classroom activities, closely resembling the construct of Behavioral Engagement (BE; Fredricks and McColskey, 2012). A child scoring high on these items likely is motivated and cognitively interacting with classroom activities, both of which are key aspects of engagement in the classroom (Guthrie et al., 2004). Distinguishing the items of BE

as separate from ATL, AF, and IC was also supported by the BE factor's relatively weak correlations to the other constructs in addition to the fit of the MMOD. Furthermore, this 3-item BE factor was present in all of the five candidate factor structures, all of which had MMOD fit statistics that outperformed the theoretical structure as measured by AIC, BIC, RMSEA, and CFI/TLI. This finding points to BE as a key feature of the dynamic underlying developmental processes related to self-regulation in the classroom.

Given that BE has been shown to relate to classroom quality and mediate reading achievement (Ponitz et al., 2009), we would expect the BE factor to uniquely predict academic achievement on the ECLS-K:2011 as well. Using a separate validation subset of the ECLS-K:2011, we found this to be the case for both reading and mathematics achievement, even in the presence of the other three self-regulation scales in the MMOD-selected factor structure. If any of these scales were collinear with BE, one would expect the presence of one to suppress the other. Instead, the fixed effect of AF as measured by the theoretical structure split evenly between the BE and AF scales of the MMOD-selected factor structure, supporting the hypothesis that the items making up those factors are measuring distinct underlying developmental processes.

Even though the overall statistical fit of the model predicting reading achievement was significantly improved by using the MMOD-selected factor structure, this was not found to be the case for mathematics achievement. This is not a surprise given that two out of the three items directly referenced the child's interactions with books. Additionally, the three BE items did not probe what cognitive strategies the child was employing, a feature of behavioral engagement that could have a differential impact on reading and mathematics. Future studies should consider adding items related to cognitive strategies as a potential way to refine and augment the predictive power of this new BE factor.

Still, even when predicting mathematics achievement, the MMOD-selected factor structure appears to reduce noise by capturing the underlying developmental processes more directly. In the MMOD-selected factor structure, all factors significantly predict mathematics achievement; in the theoretical 3-factor structure, only ATL and AF are significant predictors, suppressing IC. In general, adopting measurement structures more reflective of the underlying developmental processes allows researchers to make more informed decisions when designing interventions to manipulate outcomes. In this case, a child's behavioral engagement in the classroom emerges as a distinct construct that may provide a path to increase academic achievement and should potentially be considered by education researchers designing interventions and curricula for the classroom.

As with many important constructs in educational research, definitions of self-regulation and its underlying components vary based on perspective and approach. The MMOD allows researchers to critically examine operationalizations of a construct to test if their theoretical structure matches the underlying dynamics of the collected data. By pooling items from multiple measurement instruments in a longitudinal study and examining their factor structure and temporal dynamics using the MMOD, we were able to differentiate a BE factor as separate from ATL, AF, and IC. This new, MMOD-selected factor structure was validated by its ability to more clearly predict academic achievement on a gold-standard holdout subset of the ECLS-K:2011, further supporting the MMOD-selected factor structure as a superior representation of the underlying developmental processes surrounding self-regulation.

Limitations and Future Work

This study only used publicly available survey items from the ECLS-K:2011. It is possible that the addition of more items would have created better distinctions between constructs in our MMOD-selected factor structures. A clear indicator that our factor structures could be improved is our finding that IC negatively predicted both reading and mathematics achievement, contradicting the wealth of previous work showing IC to be positively associated with academic achievement (Blair & Raver, 2015; Blair & Razza, 2007; McClelland et al., 2015). When we omit ATL from our model, IC positively predicts reading and mathematics achievement, as expected. This suggests that our IC factor is likely measuring IC and something else that is negatively associated with academic achievement. It is possible, for example, that this confounding factor may be something like a reverse-scored indicator of a child's intellectual curiosity, which may overlap with inhibitory control but have the opposite impact on academic achievement (von Stumm et al., 2011). Unfortunately, without additional survey items related to intellectual curiosity this cannot be determined with the present data.

Future iterations of the ECLS-K:2011 may provide insight as to what this confounding factor might be by adding curiosity-related items to the ECLS-K:2011 item pool to see if it can help clarify the IC factor. It is also possible that there are other constructs related to self-regulation that would have become visible had we had a larger item pool. The items that formed our BE factor, for example, may be related to construct of "flow state", which also may impact literacy outcomes (Csikszentmihalyi, 1990; Jackson & Marsh, 1996). Augmenting the item pool with more BE and flow-related items would help illuminate the distinctions between BE and flow state, and how these processes uniquely impact academic and behavioral outcomes.

Because we only used publicly-available items in this study, a clear next step would be to expand the item pool to include restricted items collected by the ECLS-K:2011, like teacher reported measures of child behavior from the Social Skills Rating System (SSRS; Gresham and Elliott, 1990). In general, when developing and refining new scales and measurement instruments, the MMOD should be used from the beginning to insure that constructs are being considered in a longitudinal, developmental context.

Conclusion

As educational research continues to feature more longitudinal study designs, it is imperative that we critically examine the structure and temporal dynamics of the measurement models that we use in our analyses. Otherwise, we risk conflating important underlying processes or muddying the waters with redundant descriptions of identical underlying processes. The MMOD allows educational researchers to evaluate and refine measurement structures in longitudinal contexts, whether they are evaluating existing measurement instruments or working to develop new conceptualizations of underlying developmental processes. Closely matching measurement models to underlying developmental processes clarifies theory and illuminates points of leverage that can be targeted in classroom-based interventions.

References

- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4(0), 40–79. <https://doi.org/10.1214/09-SS054>
- Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., Blair, C., Nelson, K. E., & Gill, S. (2008). Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program. *Child Development*, 79(6), 1802–1817. <https://doi.org/10.1111/j.1467-8624.2008.01227.x>
- Blair, C., & Raver, C. C. (2015). School Readiness and Self-Regulation: A Developmental Psychobiological Approach. *Annual Review of Psychology*, 66(1), 711–731. <https://doi.org/10.1146/annurev-psych-010814-015221>
- Blair, C., & Razza, R. P. (2007). Relating Effortful Control, Executive Function, and False Belief Understanding to Emerging Math and Literacy Ability in Kindergarten. *Child Development*, 78(2), 647–663. <https://doi.org/10.1111/j.1467-8624.2007.01019.x>
- Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review*, 28(4), 415–427. <https://doi.org/10.1016/j.econedurev.2008.09.003>
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment*, 7(3), 309.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3), 297–334.
- Csikszentmihalyi, M. (1990). Literacy and Intrinsic Motivation. *Daedalus*, 119(2), 115–140. Retrieved March 27, 2020, from <https://www.jstor.org/stable/20025303>
- Deboeck, P. R. (2010). Estimating Dynamical Systems: Derivative Estimation Hints From Sir Ronald A. Fisher. *Multivariate Behavioral Research*, 45(4), 725–745. <https://doi.org/10.1080/00273171.2010.498294>
- Deboeck, P. R., Nicholson, J., Kouros, C., Little, T. D., & Garber, J. (2015). Integrating Developmental Theory and Methodology: Using Derivatives to Articulate Change Theories, Models, and Inferences. *Applied Developmental Science*, 19(4), 217–231. <https://doi.org/10.1080/10888691.2015.1021924>
- DiPerna, J. C., Lei, P.-W., & Reid, E. E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the Early Childhood Longitudinal Study–Kindergarten cohort. *Journal of Educational Psychology*, 99(2), 369–379. <https://doi.org/10.1037/0022-0663.99.2.369>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- Estabrook, R. (2015). Evaluating Measurement of Dynamic Constructs: Defining a Measurement Model of Derivatives. *Psychological methods*, 20(1), 117–141. <https://doi.org/10.1037/a0034523>
- Fredricks, J. A., & McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments

- (S. L. Christenson, A. L. Reschly, & C. Wylie, Eds.). In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement*. Boston, MA, Springer US. https://doi.org/10.1007/978-1-4614-2018-7_37
- Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system: Manual*. American Guidance Service.
- Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., Scafiddi, N. T., & Tonks, S. (2004). Increasing Reading Comprehension and Engagement Through Concept-Oriented Reading Instruction. *Journal of Educational Psychology*, 96(3), 403–423. <https://doi.org/10.1037/0022-0663.96.3.403>
- Hendrickson, A. E., & White, P. O. (1964). Promax: A Quick Method for Rotation to Oblique Simple Structure. *British Journal of Statistical Psychology*, 17(1), 65–70. <https://doi.org/10.1111/j.2044-8317.1964.tb00244.x>
- Husmann, K. D. (2020). Eclsk-mmod [Computer Software]. Zenodo. <https://doi.org/10.5281/zenodo.3814734>
- Husmann, K. D., Brick, T. R., & Estabrook, R. (2020). Mxmmmod: Measurement Model of Derivatives in 'OpenMx'. Retrieved March 28, 2020, from <https://CRAN.R-project.org/package=mxmmmod>
- Jackson, S. A., & Marsh, H. W. (1996). Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *Journal of Sport and Exercise Psychology*, 18(1), 17–35. <https://doi.org/10.1123/jsep.18.1.17>
- Jöreskog, K. (1969). A General Approach to Confirmatory Factor Analysis. *Psychometrika*, 34, 183–202. <https://doi.org/10.1007/BF02289343>
- Köster, J., & Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19), 2520–2522. <https://doi.org/10.1093/bioinformatics/bts480>
- Li-Grining, C. P., Votruba-Drzal, E., Maldonado-Carreño, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology*, 46(5), 1062–1077. <https://doi.org/10.1037/a0020066>
- Loevinger, J. (1957). Objective Tests as Instruments of Psychological Theory. *Psychological Reports*, 3(3), 635–694. <https://doi.org/10.2466/pr0.1957.3.3.635>
- McClelland, M. M., & Cameron, C. E. (2011). Self-regulation and academic achievement in elementary school children. *New Directions for Child and Adolescent Development*, 2011(133), 29–44. <https://doi.org/10.1002/cd.302>
- McClelland, M. M., & Cameron, C. E. (2012). Self-Regulation in Early Childhood: Improving Conceptual Clarity and Developing Ecologically Valid Measures. *Child Development Perspectives*, 6(2), 136–142. <https://doi.org/10.1111/j.1750-8606.2011.00191.x>
- McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, 43(4), 947–959. <https://doi.org/10.1037/0012-1649.43.4.947>
- McClelland, M. M., John Geldhof, G., Cameron, C. E., & Wanless, S. B. (2015). Development and Self-Regulation (R. M. Lerner, Ed.). In R. M. Lerner (Ed.), *Handbook of Child Psychology and Developmental Science*. Hoboken, NJ, USA, John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118963418.childpsy114>
- Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: In-*

- terdisciplinary Research & Perspective*, 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1
- Molenaar, P. C. M., De Gooijer, J. G., & Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika*, 57(3), 333–349. <https://doi.org/10.1007/BF02295422>
- Najarian, M., Tourangeau, K., Nord, C., & Wallner-Allen, K. (2017). Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K: 2011), Kindergarten Psychometric Report. *National Center for Education Statistics, Institute of Education Sciences, US Department of Education. Washington, DC, NCES 2015-078*.
- Najarian, M., Tourangeau, K., Nord, C., & Wallner-Allen, K. (2018). Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K: 2011), First-Grade and Second-Grade Psychometric Report. *National Center for Education Statistics, Institute of Education Sciences, US Department of Education. Washington, DC, NCES 2018-183*.
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*, 81(2), 535–549. <https://doi.org/10.1007/s11336-014-9435-8>
- Pan, Q., & Zhu, Q. (2018). Development of self-control in early childhood—a growth mixture modeling approach (S. Elmer, Ed.). *Cogent Psychology*, 5(1), 1544537. <https://doi.org/10.1080/23311908.2018.1544537>
- Ponitz, C. C., Rimm-Kaufman, S. E., Grimm, K. J., & Curby, T. W. (2009). Kindergarten Classroom Quality, Behavioral Engagement, and Reading Achievement. *School Psychology Review*, 38(1), 20.
- Putnam, S. P., & Rothbart, M. K. (2006). Development of Short and Very Short Forms of the Children’s Behavior Questionnaire. *Journal of Personality Assessment*, 87(1), 102–112. https://doi.org/10.1207/s15327752jpa8701_09
- R Core Team. (2019). *R: A language and environment for statistical computing, version 3.0.2. Vienna, Austria: R Foundation for Statistical Computing; 2013*.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, 48(2), 1–36.
- Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of Temperament at Three to Seven Years: The Children’s Behavior Questionnaire. *Child Development*, 72(5), 1394–1408. <https://doi.org/10.1111/1467-8624.00355>
- Sung, J., & Wickrama, K. A. (2018). Longitudinal relationship between early academic achievement and executive function: Mediating role of approaches to learning. *Contemporary Educational Psychology*, 54, 171–183. <https://doi.org/10.1016/j.cedpsych.2018.06.010>
- Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., & Najarian, M. (2009). Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K): Combined User’s Manual for the ECLS-K Eighth-Grade and K–8 Full Sample Data Files and Electronic Codebooks. *National Center for Education Statistics, NCES 2009-004*.
- Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L., & Najarian, M. (2019). Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011): User’s Manual for the ECLS-K:2011 Kindergarten–Fifth

- Grade Data File and Electronic Codebook, Public Version. *U.S. Department of Education. Washington, DC: National Center for Education Statistics, NCES 2019-051*, 347.
- Tripodis, Y., & Ziropiannis, N. (2015). Dynamic Factor Analysis for Multivariate Time Series: An Application to Cognitive Trajectories. *International journal of clinical biostatistics and biometrics*, 1(1). Retrieved June 19, 2019, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4704801/>
- von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The Hungry Mind: Intellectual Curiosity Is the Third Pillar of Academic Performance. *Perspectives on Psychological Science*, 6(6), 574–588. <https://doi.org/10.1177/1745691611421204>
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial Invariance within Longitudinal Structural Equation Models: Measuring the Same Construct across Time. *Child development perspectives*, 4(1), 10–18. <https://doi.org/10.1111/j.1750-8606.2009.00110.x>

Appendix

Candidate Factor Structures

Table A1

2-Factor Candidate Structures

Theoretical scale	Measure	Fall of Kindergarten		Spring of Kindergarten		Spring of 1st Grade	
		ATL+AF+IC	BE	ATL+AF+IC	BE	ATL+AF+IC	BE
ATL	TKEEPS	0.634	0.075	0.677	0.057	0.686	-0.010
	TSHOWS	0.457	0.262	0.458	0.254	0.548	0.144
	TWORKS	0.613	0.209	0.647	0.190	0.750	0.056
	TADAPTS	0.702	-0.056	0.657	0.023	0.706	-0.082
	TFOLLOW	0.899	-0.126	0.877	-0.099	0.869	-0.109
	TPERSIS	0.546	0.248	0.602	0.206	0.731	0.105
	TATTEN	0.839	0.066	0.844	0.059	0.898	-0.012
AF	TBEZDAC	-0.726	-0.134	-0.747	-0.097	-0.809	-0.054
	TBNOFIN	-0.638	-0.158	-0.729	-0.046	-0.757	-0.050
	TBGCCLR	0.086	0.759	0.132	0.679	0.142	0.676
	TBGCBLD	-0.132	0.915	-0.081	0.841	-0.030	0.850
	TBEZDSL	-0.720	-0.086	-0.719	-0.126	-0.781	-0.042
	TBABS BK	-0.002	0.626	-0.067	0.711	-0.076	0.559
IC	TBWTTSK	0.624	0.065	0.655	-0.006	0.525	0.118
	TBTRBST	-0.824	0.077	-0.808	0.049	-0.800	0.047
	TBFLWIN	0.750	0.094	0.827	0.033	0.827	0.034
	TBSTNO	0.830	-0.101	0.882	-0.168	0.797	-0.116

Note: Items in bold are the factor loadings with the highest absolute value across factors and included in the factor parcel.

Table A2*2-Factor Candidate Structure Intercorrelations: Fall of Kindergarten*

	ATL+AF+IC	BE
ATL+AF+IC	1.000	
BE	0.733	1.000

Table A3*2-Factor Candidate Structure Intercorrelations: Spring of Kindergarten*

	ATL+AF+IC	BE
ATL+AF+IC	1.000	
BE	0.702	1.000

Table A4*2-Factor Candidate Structure Intercorrelations: Spring of 1st Grade*

	ATL+AF+IC	BE
ATL+AF+IC	1.000	
BE	0.583	1.000

Table A5*3-Factor Candidate Structures*

Theoretical scale	Measure	Fall of Kindergarten			Spring of Kindergarten			Spring of 1st Grade		
		ATL	AF+IC	BE	ATL	AF+IC	BE	ATL	AF+IC	BE
ATL	TKEEPS	0.482	0.254	-0.004	0.372	0.375	0.013	0.361	0.362	-0.013
	TSHOWS	0.784	-0.143	0.104	0.775	-0.133	0.080	0.769	-0.141	0.079
	TWORKS	0.791	0.008	0.053	0.926	-0.062	-0.011	0.889	-0.041	-0.019
	TADAPTS	0.613	0.218	-0.164	0.488	0.265	-0.054	0.440	0.310	-0.094
	TFOLLOW	0.237	0.684	-0.114	0.197	0.691	-0.069	0.144	0.751	-0.075
	TPERSIS	0.752	-0.029	0.099	0.799	-0.010	0.034	0.869	-0.044	0.036
AF	TATTEN	0.453	0.474	0.011	0.553	0.399	-0.020	0.586	0.372	-0.031
	TBEZDAC	-0.246	-0.530	-0.117	-0.426	-0.405	-0.043	-0.580	-0.291	-0.028
	TBNOFIN	-0.321	-0.388	-0.116	-0.452	-0.363	0.014	-0.590	-0.229	-0.019
	TBGCCLR	0.047	0.103	0.726	0.071	0.126	0.646	0.034	0.137	0.678
	TBGCBLD	0.039	-0.085	0.854	-0.070	0.007	0.867	0.061	-0.044	0.816
	TBEZDSL	-0.031	-0.694	-0.120	-0.179	-0.571	-0.135	-0.318	-0.501	-0.051
IC	TBABS BK	-0.019	0.058	0.609	0.072	-0.056	0.645	-0.016	-0.038	0.547
	TBWTTSK	-0.006	0.624	0.104	-0.092	0.714	0.074	-0.101	0.632	0.167
	TBTRBST	0.174	-0.967	0.002	0.079	-0.856	-0.038	-0.032	-0.789	0.005
	TBFLWIN	0.268	0.532	0.077	0.231	0.627	0.043	0.267	0.595	0.051
	TBSTNO	0.003	0.813	-0.049	-0.120	0.961	-0.072	-0.130	0.948	-0.069

Note: Items in bold are the factor loadings with the highest absolute value across factors and included in the factor parcel.

Table A6*3-Factor Candidate Structure Intercorrelations: Fall of Kindergarten*

	ATL	AF+IC	BE
ATL	1.000		
AF+IC	0.802	1.000	
BE	0.679	0.662	1.000

Table A7*3-Factor Candidate Structure Intercorrelations: Spring of Kindergarten*

	ATL	AF+IC	BE
ATL	1.000		
AF+IC	0.801	1.000	
BE	0.662	0.619	1.000

Table A8*3-Factor Candidate Structure Intercorrelations: Spring of 1st Grade*

	ATL	AF+IC	BE
ATL	1.000		
AF+IC	0.793	1.000	
BE	0.543	0.509	1.000

Table A9

4-Factor Candidate Structures

Theoretical scale	Measure	Fall of Kindergarten				Spring of Kindergarten				Spring of 1st Grade			
		ATL	AF	IC	BE	ATL	AF	IC	BE	ATL	AF	IC	BE
ATL	TKEEPS	0.502	0.128	0.120	-0.005	0.347	0.144	0.279	0.019	0.248	0.277	0.228	-0.013
	TSHOWS	0.793	-0.141	-0.028	0.111	0.771	-0.094	-0.032	0.093	0.763	-0.071	-0.037	0.087
	TWORKS	0.812	0.105	-0.112	0.030	0.829	0.127	-0.080	-0.006	0.753	0.159	-0.034	-0.013
	TADAPTS	0.658	-0.151	0.321	-0.131	0.544	-0.124	0.347	-0.036	0.541	-0.164	0.417	-0.091
	TFOLLOW	0.260	0.230	0.462	-0.090	0.239	0.022	0.657	-0.050	0.154	0.128	0.653	-0.069
	TPERSIS	0.767	0.039	-0.082	0.084	0.726	0.110	-0.032	0.037	0.740	0.149	-0.034	0.041
	TATTEN	0.477	0.382	0.108	-0.013	0.458	0.338	0.189	-0.025	0.404	0.392	0.198	-0.031
AF	TBEZDAC	-0.258	-0.660	0.064	-0.048	-0.227	-0.776	0.091	0.014	-0.244	-0.754	0.080	-0.006
	TBNOFIN	-0.340	-0.412	-0.006	-0.075	-0.300	-0.579	0.007	0.051	-0.307	-0.606	0.055	-0.005
	TBGCCLR	0.065	0.128	0.008	0.687	0.056	0.126	0.043	0.628	-0.002	0.110	0.078	0.672
	TBGCBLD	0.043	-0.113	0.036	0.861	-0.038	-0.056	0.039	0.869	0.039	0.034	-0.045	0.808
	TBEZDSL	0.000	-0.861	0.037	-0.040	0.016	-0.730	-0.103	-0.099	-0.023	-0.691	-0.148	-0.035
	TBABS BK	-0.014	0.034	0.050	0.593	0.071	0.004	-0.045	0.635	0.009	-0.060	-0.001	0.548
IC	TBWTTSK	-0.040	-0.031	0.714	0.174	-0.027	-0.036	0.706	0.099	-0.014	-0.048	0.619	0.179
	TBTRBST	0.190	-0.872	-0.214	0.061	0.235	-0.682	-0.398	-0.007	0.236	-0.685	-0.418	0.021
	TBFLWIN	0.289	0.198	0.345	0.089	0.220	0.202	0.469	0.049	0.231	0.184	0.484	0.056
	TBSTNO	-0.029	0.108	0.790	-0.003	-0.071	0.065	0.886	-0.053	-0.063	0.064	0.860	-0.062

Note: Items in bold are the factor loadings with the highest absolute value across factors and included in the factor parcel.

Table A10*4-Factor Candidate Structure Intercorrelations: Fall of Kindergarten*

	ATL	AF	IC	BE
ATL	1.000			
AF	0.785	1.000		
IC	0.728	0.742	1.000	
BE	0.667	0.647	0.548	1.000

Table A11*4-Factor Candidate Structure Intercorrelations: Spring of Kindergarten*

	ATL	AF	IC	BE
ATL	1.000			
AF	0.760	1.000		
IC	0.734	0.764	1.000	
BE	0.637	0.615	0.563	1.000

Table A12*4-Factor Candidate Structure Intercorrelations: Spring of 1st Grade*

	ATL	AF	IC	BE
ATL	1.000			
AF	0.752	1.000		
IC	0.709	0.746	1.000	
BE	0.515	0.510	0.468	1.000

Table A13

5-Factor Candidate Structures

Theoretical scale	Measure	Fall of Kindergarten					Spring of Kindergarten					Spring of 1st Grade				
		ATL	AF	AF2	IC	BE	ATL	AF	AF2	IC	BE	ATL	AF	AF2	IC	BE
ATL	TKEEPS	0.466	0.094	0.081	0.125	-0.006	0.360	0.051	0.105	0.261	0.018	0.163	0.159	0.244	0.266	-0.022
	TSHOWS	0.851	-0.120	-0.062	-0.075	0.131	0.844	-0.093	-0.034	-0.081	0.080	0.853	-0.015	-0.141	-0.073	0.098
	TWORKS	0.788	0.060	0.085	-0.123	0.039	0.817	0.151	-0.018	-0.069	-0.002	0.694	0.120	0.125	-0.016	-0.011
	TADAPTS	0.625	0.032	-0.149	0.322	-0.132	0.555	-0.028	-0.105	0.334	-0.036	0.494	-0.145	0.000	0.438	-0.092
	TFOLLOW	0.329	-0.154	0.317	0.425	-0.068	0.306	-0.150	0.157	0.599	-0.062	0.180	0.160	-0.083	0.654	-0.062
	TPERSIS	0.688	0.199	-0.063	-0.059	0.071	0.690	0.211	-0.089	-0.000	0.045	0.653	0.068	0.209	-0.005	0.036
	TATTEN	0.541	-0.088	0.443	0.049	0.011	0.543	-0.019	0.370	0.113	-0.043	0.425	0.382	0.023	0.184	-0.023
AF	TBEZDAC	-0.130	-0.424	-0.421	0.025	-0.017	-0.162	-0.483	-0.364	0.044	-0.003	-0.206	-0.600	-0.286	0.070	-0.007
	TBNOFIN	-0.111	-0.680	-0.066	-0.078	0.005	-0.144	-0.689	-0.074	-0.053	0.045	-0.187	-0.434	-0.443	0.048	0.014
	TBGCCLR	0.015	0.154	0.034	0.036	0.672	0.013	0.142	-0.002	0.084	0.638	-0.062	0.028	0.168	0.105	0.672
	TBGCBLD	0.048	-0.011	-0.099	0.034	0.853	-0.064	0.030	-0.086	0.074	0.864	0.026	0.013	0.063	-0.036	0.789
	TBEZDSL	-0.033	-0.049	-0.828	0.084	-0.062	-0.042	-0.070	-0.764	0.010	-0.082	-0.082	-0.726	0.010	-0.086	-0.048
	TBABS BK	0.023	-0.080	0.081	0.026	0.608	0.125	-0.145	0.129	-0.089	0.634	0.050	-0.009	-0.104	-0.019	0.559
IC	TBWTTSK	-0.035	-0.020	-0.017	0.711	0.168	-0.046	0.004	-0.040	0.722	0.105	-0.032	-0.046	-0.008	0.637	0.175
	TBTRBST	0.147	-0.006	-0.860	-0.173	0.038	0.183	-0.031	-0.728	-0.313	0.007	0.178	-0.659	-0.015	-0.390	0.012
	TBFLWIN	0.262	0.093	0.143	0.352	0.085	0.202	0.133	0.090	0.479	0.056	0.214	0.160	0.045	0.496	0.057
	TBSTNO	-0.079	0.096	0.032	0.846	-0.019	-0.098	0.037	0.030	0.913	-0.045	-0.093	0.034	0.025	0.898	-0.067

Note: Items in bold are the factor loadings with the highest absolute value across factors and included in the factor parcel.

Table A14*5-Factor Candidate Structure Intercorrelations: Fall of Kindergarten*

	ATL	AF	AF2	IC	BE
ATL	1.000				
AF	0.710	1.000			
AF2	0.757	0.697	1.000		
IC	0.738	0.607	0.747	1.000	
BE	0.655	0.596	0.618	0.554	1.000

Table A15*5-Factor Candidate Structure Intercorrelations: Spring of Kindergarten*

	ATL	AF	AF2	IC	BE
ATL	1.000				
AF	0.698	1.000			
AF2	0.738	0.695	1.000		
IC	0.746	0.658	0.749	1.000	
BE	0.642	0.535	0.592	0.553	1.000

Table A16*5-Factor Candidate Structure Intercorrelations: Spring of 1st Grade*

	ATL	AF	AF2	IC	BE
ATL	1.000				
AF	0.696	1.000			
AF2	0.576	0.470	1.000		
IC	0.710	0.733	0.541	1.000	
BE	0.500	0.476	0.374	0.467	1.000