Hammill Institute on Disabilities

Remedial and Special Education
2018, Vol. 39(5) 274–288

DOI: 10.1177/0741932518772904
rase.sagepub.com
SAGE

Extending the Simple View of Reading to Account for Variation Within Readers and Across Texts: The Complete View of Reading (CVR*i*)

David J. Francis, PhD[1], Paulina A. Kulesz, PhD[1], and Julia S. Benoit, PhD[1]

Abstract

This study leverages advances in multivariate cross-classified random effects models to extend the Simple View of Reading to account for variation within readers and across texts, allowing for both the personalization of the reading function and the integration of the component skills and text and discourse frameworks for reading research. We illustrate the Complete View of Reading (CVR*i*) using data from an intensive longitudinal design study with a large sample of typical (*N* = 648) and struggling readers (*N* = 865) in middle school and using oral reading fluency as a proxy for comprehension. To illustrate the utility of the CVR*i*, we present a model with cross-classified random intercepts for students and passages and random slopes for growth, Lexile difficulty, and expository text type at the student level. We highlight differences between typical and struggling readers and differences across students in different grades. The model illustrates that readers develop differently and approach the reading task differently, showing differential impact of text features on their fluency. To be complete, a model of reading must be able to reflect this heterogeneity at the person and passage level, and the CVR*i* is a step in that direction. Implications for reading interventions and 21st century reading research in the era of "Big Data" and interest in phenotypic characterization are discussed.

Keywords

Simple View of Reading, reading comprehension, oral reading fluency, multilevel models, cross-classified random effects models

Introduction

Reading Comprehension Frameworks

Reading comprehension has been viewed as the end product of complex interactions between a reader, text, and activity. The quality of this product depends on the skills that the reader brings to the task, the demands that the text places on the reader, and the challenge posed by the specific activity in which the reader is engaged (Snow, 2002). The complex interactive relations that are involved in constructing a coherent mental representation of the text are rarely fully integrated in reading research. During any period in the history of research on reading comprehension, one finds several lines of active research that focus on different aspects of comprehension. Specifically, we find that reading research can be characterized by its focus on the component skills of readers, the features of text and linguistic discourse that affect comprehension, and the development of reading as children move from the early stages of reading acquisition to reading mastery. For ease of communication, we think of these lines of research as frameworks for studying reading and have found utility in labeling them the Component Skills Framework (CSF), the Text and Discourse Framework (TDF), and the Developmental Framework (DevF), respectively. The lines of research pursued by these frameworks may cross and interact somewhat, but hardly ever are they fully integrated.

The Simple View of Reading (SVR; Gough & Tunmer, 1986) has contributed greatly to our understanding of reading, both as a heuristic model for conceptualizing the process of comprehending written language and also as a framework for reading research. The CSF of reading research is a direct descendant of the SVR. This framework

[1]University of Houston, TX, USA

Corresponding Author:
David J. Francis, Department of Psychology, Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, 4849 Calhoun Road, Room 372, Houston, TX 77204-6022, USA.
Email: dfrancis@uh.edu

has sought to elaborate on the component linguistic and non-linguistic cognitive skills that underlie the two major components of the SVR—decoding and language comprehension. Since the SVR was originally posited, investigators have elaborated on these component skills and have shifted their attention toward deeper comprehension processes that are necessary for building a coherent text representation. The TDF departs significantly from the CSF and the SVR by focusing on the role of text characteristics and the features of linguistic discourse and how these aspects of language affect comprehension. The TDF has underscored the importance of helping struggling students to recognize different features and styles of text, and to understand how text is organized to achieve an author's goals.

While somewhat of an oversimplification, it is not wholly unfair to characterize the CSF as a reading framework that is focused on the reader and the TDF as a framework that is focused on the text. When one examines the research on reading, CSF research has focused on the impact of different cognitive and linguistic skills of the reader on comprehension, as well as demographic features of readers (e.g., good vs. poor, typically developing vs. disabled, female vs. male, and English Language Learner [ELL] vs. native speaker), and to a lesser extent motivational and attitudinal characteristics of readers (see Ahmed et al., 2016; Cain & Oakhill, 2009; Joshi & Aaron, 2000; Nation & Snowling, 2004; Oliviera, da Silva, Dias, Seabra, & Macedo, 2014, as recent examples of CSF research). Interventions in the CSF have focused on building stronger cognitive and linguistic skills in the reader, or improving their motivation and interest in reading (Scammacca, Roberts, Vaughn, & Stuebing, 2015). In contrast, and again oversimplifying, research conducted in the TDF has typically focused on the text and the features of the text and how these influence comprehension, and to a much lesser extent on the cognitive, linguistic, and motivational characteristics of the reader that have been the focus of the CSF (Lynch & van den Broek, 2007; McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996). Interventions in the TDF framework have sought to engage readers in identifying and navigating the discourse features of the text and on building argumentation skills in the reader (Meyer & Ray, 2011; Pyle et al., 2017). In addition to these two broad frameworks of reading research, a third organizing framework for reading research is the DevF, which has primarily focused on the effects of developmental changes in reading, both in the reader characteristics and the discourse level skills that affect comprehension (Barnes, Dennis, & Haefele-Kalvaitis, 1996; Cain, Oakhill, & Bryant, 2004; Hannon & Daneman, 2009; Lynch & van den Broek, 2007; McNamara, Graesser, & Louwerse, 2012; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; van den Broek, Lorch, & Thurlow, 1996; Wagner, Torgesen, & Rashotte, 1994).

On their own, each of these frameworks is powerful and useful and has led to significant advances in the science of reading and reading comprehension. To achieve these gains, research conducted under a given framework tends to either average over, or to control variation that is the primary focus of research in other frameworks. For example, it is not uncommon for research on the role of component skills to treat texts as a sampling feature and to administer measures of reading achievement that average over different texts, or to control text features through the use of readability formulae. Research conducted in the TDF may focus on typically developing readers in a narrow grade span, and then average over readers to assess the effects of text features on students' understanding. These approaches are valuable in that they serve to isolate the contributions of different elements of the reader–text-purpose interaction, either at a given time point or developmentally. However, individually they fail to address the potential interactions that exist across elements and they presume that the effects observed between individuals also apply within individuals. Inferring that a relation observed at one level of analysis applies at a different level of analysis has been termed the ecological fallacy (Francis, Barr, Benoit, & McIntyre, 2017) in multilevel modeling research. We propose that it is both possible and desirable to develop models of reading that are not susceptible to the ecological fallacy. Specifically, we propose a model of reading that is capable of capturing the variation that exists within readers, as well as the variation among readers, across texts and time, to best understand how particular readers function when faced with comprehending specific texts for specific purposes. These models also describe how readers generally function in these varying contexts. We refer to such models as *personalized models* of reading.

Prior Research on Integrating Reading Frameworks

To develop *personalized models* of reading requires overcoming the challenge of integrating broad literatures on component skills, text and discourse, and reading development. These literatures exist largely in parallel. The joint investigation of reader and text characteristics has been limited in three ways: (a) by limiting the focus to only one or two of the three domains of factors (reader characteristics, text characteristics, and/or comprehension activity) and averaging across the other domain(s), (b) by limiting the focus within each domain to a few features whenever more than one domain is considered, and (c) by generally failing to consider how reader skills and text features further interact with reader activity.

Prior research has investigated the effects of reader and text characteristics on comprehension, but has generally failed to systematically study both component skills and text characteristics and interactions with different kinds of comprehension processes (see, for example, Best, Floyd, & McNamara, 2008; Garcia & Cain, 2014; Millis, Graesser, & Haberlandt, 1993; Yap, Tse, & Balota, 2009). Although these studies examined relations between some reader characteristics and some text characteristics, a number of other characteristics were not examined, and in no case did the studies consider how the relations might have varied across different kinds of comprehension items (i.e., the reader activity), such as remembering text information versus making text inferences, or integrating text information with background knowledge. Generally, individual studies have examined a limited spectrum of reader and text characteristics, which, in turn, has limited the number of reader–text interactions that could be examined in any particular study. For example, McNamara and colleagues (1996) found that high coherence in texts afforded greater benefits to low-knowledge readers and low coherence texts afforded greater benefit to high-knowledge readers, while Ozuru, Dempsey, and McNamara (2009) found that cohesion affects answering text memory questions but not text inference questions. Each study considered only a subset of possible reader–text interactions, and neither study considered whether distinct aspects of cohesion (i.e., referential and deep cohesion) would yield comparable results. Because these two types of cohesion reflect different text characteristics, there is merit in keeping them distinct when studying the effects of reader–text interactions on comprehension.

While reading research has historically progressed within each framework with limited integration across frameworks, Kulesz, Francis, Barnes, and Fletcher (2016) showed how the TDF and CSF could be integrated using advances in multilevel modeling and item-response modeling. Specifically, Kulesz and colleagues successfully integrated these two frameworks to simultaneously study the roles of component skills and text characteristics and their interactions using explanatory item-response models (E-item response theory [IRT]). E-IRT models are a type of cross-classified random effects model for binary or multinomial outcomes that simultaneously model random variation on the person side (i.e., ability) and on the item side (i.e., item difficulty) of the model. These models are most easily understood as multilevel models where the levels are crossed rather than hierarchically nested. The present study employs cross-classified random effects models for continuously measured outcomes to integrate the CSF and TDF in a developmental model of oral reading fluency (ORF), extending the work of Kulesz et al. (2016) to continuous outcomes and the work of Barth, Tolar, Fletcher, and Francis (2014) to longitudinal contexts.

Limitations of the SVR

Despite its many strengths and phenomenal record as a guiding framework for reading research, the SVR is limited in several ways, not all of equal importance. First and foremost, the SVR is deterministic in that it lacks a stochastic component, that is, the model lacks a random error component. The omission of a stochastic component reflects the fact that the authors viewed the SVR as more of a conceptual model than a statistical model, which they formulated as multiplicative to highlight the fact that without decoding skill or linguistic comprehension there can be no comprehension of written language. It is perhaps noteworthy that most investigations of the SVR have ignored both its multiplicative aspect and its failure to include a stochastic component. Adding a stochastic component to the SVR makes it better suited to function as a statistical model.

As originally formulated, the SVR is not explicitly developmental in nature. In fact, as originally formulated, the SVR captures interindividual differences in reading. That is, the SVR is a framework for understanding differences between individuals in comprehension at a given point in time. Some research involving the SVR has focused explicitly on developmental changes in the roles of different component skills that underlie the decoding and language comprehension components of the SVR (Gough, Hoover, & Peterson, 1996; Schatschneider et al., 2004; Tighe & Schatschneider, 2014; Verhoeven & Van Leeuwe, 2008). However, even in these studies, the focus of any given modeling effort has been to describe individual differences in comprehension on the basis of differences between individuals in the component skills (Mehta, Foorman, Branum-Martin, & Taylor, 2005). The SVR can be made explicitly developmental without much difficulty by simply adding a time subscript to the components of the model (Decoding and Language Comprehension) and the outcome (Reading Comprehension). However, such a simple modification is itself limiting because subscripting only the elements of the SVR for time implies both that the function that relates the components to comprehension is developmentally invariant (i.e., all of the developmental variation is reflected in the values of the components and the outcome) and that the function is invariant across individuals (i.e., the components combine in precisely the same way for everyone and at all times). At the same time, developmental research conducted under the CSF suggests that the roles of various component skills change as readers mature (Catts, Hogan, & Adlof, 2005; Schatschneider et al., 2004), which is consistent with Carroll's hypothesis that human cognitive abilities evolve from a relatively undifferentiated mass into a diverse multifactor structure as individuals develop (Carroll, 1993; see Note 1). These findings and Carroll's theory of human development call into question whether simply adding time subscripts for the components and outcomes would result in a suitable developmental SVR. Allowing for developmental variation in the function that relates the components to comprehension and interindividual differences in that developmental function requires more extensive modification of the SVR.

A final limitation of the SVR is similarly challenging to overcome. Because the SVR is the quintessential model of reading comprehension crafted in the CSF, it fails to explicitly account for variation in text a la the TDF. Regardless of the intent of the original authors, the SVR is typically viewed as a latent trait model, with decoding skill and linguistic comprehension viewed as latent traits that combine multiplicatively to determine an individual's standing on the latent trait of reading comprehension. As such, the model cannot account for significant variation in an individual's comprehension across different texts at the same point in time nor can the model account for differences in comprehension between individuals that vary in magnitude significantly across different texts. (*Note.* To say that the differences vary significantly, we mean to imply only that the variation in differences is not simply the result of random error.) In short, even at a given point in time in a reader's development, comprehension is not a static latent trait but a dynamically varying product that is influenced by the component skills of the reader and the features of the text and their interactions. Certainly, it is possible to conceptualize reading comprehension as a latent trait in the sense of a reader's ability to understand written language in a general sense, which one might estimate by averaging over their performance on many texts of fixed or varying difficulty. Such a conceptualization has great utility for describing general differences between individuals, and for tracking the development of a reader's general ability to comprehend written language. But such a conceptualization of comprehension is also limited because it implies that readers' comprehension is invariant across texts of equal difficulty and that differences between individuals are static across different texts, except for measurement error. To accommodate such dynamic variation in reading comprehension within and between individuals requires substantial extension of the SVR.

*Rationale for a Complete View of Reading (CVR**i**)*

Integration of the components skills, text discourse, and developmental literatures is necessary to build personalized models of reading comprehension from a developmental perspective. Although relations between reading comprehension with reader and text characteristics have been examined in cross-sectional contexts, much less work has been done in longitudinal contexts. The current study seeks to develop a theoretical framework for personalized models of reading by examining changes in the effects of reader and text characteristics along with their interactions on ORF. In recent years, there has been growing evidence for using ORF measures as proxies for assessing reading comprehension (Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008). Prior research suggests that efficient and automatic word reading, which is described as oral reading fluency, is more closely related to reading comprehension than decoding accuracy (Good, Simmons, & Kame'enui, 2001; Roberts, Good, & Corcoran, 2005). To develop personalized models of reading, ORF serves as an excellent proxy for comprehension because measures of fluency can be collected with minimal time investment in comparison to measuring reading comprehension. The challenge for the development of personalized models of reading is to require students to read a broad array of passages that vary in substance and difficulty so that variability within and between students, within and between passages, and within and across time can be modeled.

A theoretical framework for the current study builds directly on reading research that has focused on component skills models of reading, on characteristics of text and discourse that influence readers' ability to understand written language, and on developmental changes in the roles of different component skills and text features on comprehension. Based on the component skills literature, we know the following: (a) accurate and fluent word reading warrants meaningful and efficient processing of words and sentences (Perfetti & Stafura, 2014), (b) vocabulary and background knowledge (world and domain-specific knowledge) are critical for reading comprehension because the two sources of knowledge help in understanding relations between words and deriving meaning of sentences (Cain & Oakhill, 2007; Nation & Snowling, 2004; Perfetti & Adolf, 2012; Vellutino, Tunmer, Jaccard, & Chen, 2007), and (c) working memory, a temporary storage of processed information, is necessary to integrate information within and between sentences (van den Broek, 2012).

Based on the discourse literature, we know that text features demarcating text difficulty (i.e., word frequency, sentence length, cohesion, and genre) affect reading comprehension, with less frequent words, longer sentences, low referential cohesion (i.e., low overlap of words and concepts across the text), low deep cohesion (i.e., less connecting words used to clarify relations between information in the text), and expository genre characterizing more difficult texts (Graesser & McNamara, 2011; McNamara, 2001; McNamara et al., 1996).

Based on the developmental literature, we know the following: (1) correlational and predictive relations between word reading skills and comprehension decrease with increasing age, (2) younger students who allocate limited working memory resources to word decoding are less effective in building coherent text representations (Gough et al., 1996; Tighe & Schatschneider, 2014; Verhoeven & Van Leeuwe, 2008), (3) even though vocabulary and background knowledge are important in predicting comprehension across a broad age range, the contributions of vocabulary and background knowledge to comprehension increase with age and text difficulty (Ahmed et al., 2016; Oakhill & Cain, 2012; Storch & Whitehurst, 2002), (4) working memory explains unique variance in reading comprehension in students of different ages (Cain et al., 2004), (5) younger readers are more sensitive to (a) texts including low frequency words because younger readers tend to have lower levels of depth and breadth of vocabulary (van den Broek et al., 1996), (b) text cohesion because younger readers have limited text processing skills necessary to infer relations not explicitly stated in the text (Lynch & van den Broek, 2007), and (c) text genre because expository texts are differentially harder for younger readers to process (McNamara et al., 2012), and (6) younger readers tend to make less text-based and knowledge-based inferences relative to older readers because older readers infer more relations between groups of events allowing for identification of main themes in the text (Barnes et al., 1996; Hannon & Daneman, 2009).

As mentioned previously, the developmental literature has shown that the role of component skills in reading is not static (Schatschneider et al., 2004). However, prior developmental work has not simultaneously examined the changing nature of texts over development, and the extent to which the shifting roles of component skills are a function of how reading is assessed and the nature of texts used in reading comprehension assessments (Francis, Fletcher, Catts, & Tomblin, 2005), or a function of changes in the actual importance of these skills in comprehension.

This work seeks to posit a more Complete View of Reading (CVR*i*) to address the specific limitations of the SVR noted above. In addition to extending the SVR to include a stochastic component, we pursue extensions to capture both developmental variations within and between individuals, as well as variation in performance across texts within time, both within and between individuals. To be fair, our model is not so much an extension of the SVR as an alternative to the SVR that addresses the above limitations. We refer to the model as the CVR*i* in that it attempts to integrate research from the CSF and the TDF in an explicitly developmental statistical model that is person specific, hence the subscript *i* after CVR to denote the individualized nature of the model. We propose this model as a first step in the development of personalized models of reading, paving the way for researchers to design and test interventions that are person-specific and sensitive to developmental variation in the factors that affect reading comprehension generally (i.e., consistently across individuals) as well as in person-specific ways. We offer the CVR*i* as an attempt to integrate different reading research frameworks in a comprehensive, developmental model. To do so, we will adapt the models of Kulesz et al. (2016) to measures of ORF and the models of Barth et al. (2014) to developmental contexts.

A heuristic representation of the CVR*i* in a mathematical form is given by equation 1. The model is heuristic because we formulate the model here only in a general sense, and not explicitly as a multilevel model. We provide more detail on the specific models to be fit below.

$$\text{CVR}i \quad Y_{ipt} = f_i\left(X, D_t, Q_p, \beta, U_i, U_p, \varepsilon_{ipt}\right) \tag{1}$$

Equation 1 states that comprehension (or in this case, ORF) for person *i* reading passage *p* at time *t* is a function, given by f_i of three sets of variables, X, D_t, and Q_p, a set of regression weights, β, and three sets of random components, U_i, U_p, ε_{ipt}. The three sets of variables represent static component skills (X) of the reader (i.e., time invariant characteristics of the person, such as gender, vocabulary skill, or working memory measured at a single point in time), developmental, or dynamic characteristics of the reader (D_t), including variables such as age, months of instruction, or sessions of intervention, but also other time varying covariates, such as motivation or decoding skill measured over time, and characteristics of the passages (Q_p). The subscript *t* on D_t is provided to distinguish these person characteristics as time varying from those person characteristics in X, which are time invariant. All three variable sets (X, D_t, and Q) are matrices, the dimensions of which we leave ambiguous since the specific dimensions will depend on the study design. It is easy to see how varying the reading activity could be captured in variable sets X and D_t, depending on the specific study design context, but also how component skills of the reader may be represented either in X, if they are measured statically, or in D_t, if they are measured over time (see Note 2). The regression weights (β) capture the fixed effects associations between the variables (X, D_t, Q_p) and the response (Y_{ipt}), whereas U_i, U_p, and ε_{ipt} capture the random effects of the model, that is, the person-specific (U_i) and passage-specific (U_p) effects, and the residual error (ε_{ipt}) for person *i* on passage *p* at time *t*. Importantly, the function, f_i, is person specific as evidenced by the subscript *i*, as well as by the inclusion of the random person effects, U_i. In this heuristic, the random slopes are simply the products of stochastic components, U_i and variables of interest. For example, in a growth-modeling context with person-specific slopes, the product of D_t and β captures the common effect of age, whereas the product of D_t and U_i captures the person specific part of the slope. In our formulation in this article, f_i is linear in the parameters and explanatory variables, but f_i need not be restricted to the class of linear functions. Below, we will operationalize the heuristic description of the CVR*i* as a multivariate cross-classified random effects model for our specific design, but the CVR*i* is not restricted to this specific statistical representation, even for our particular design. For example, it is easy to see how the model could be represented as a cross-classified random effects model for binary or multinomial responses in an explanatory item-response design, or as an exponential (i.e., inherently nonlinear) growth function in an expanded developmental context. It is also possible to expand equation 1 to include latent variable model formulations, including models with latent classes, or growth mixtures (Boscardin, Muthen, Francis, & Baker, 2008; Lubke & Muthen, 2005).

Method

Subjects

This study was based on a secondary data analysis of data collected during the first year (2006–2007) of Project 1 of the Texas Center for Learning Disabilities (www.texasld-center.org). This phase of the research included assessment and longitudinal monitoring of a large cohort of students in Grades 6 to 8 (see Barth et al., 2014, for a more detailed description of the larger project). The total sample for the current analyses (n = 1,513) included 84% of the original study participants (n = 1,794) and included 648 (43%) typical readers (randomly sampled from the pool of screened typical readers), and 865 (57%) struggling readers. About half of the participants were females (51%); African American and Hispanics made up more than 75% of the study sample (African American: 40%, Hispanic: 36%, White: 20%, and Other: 4%); and nearly 60% were on free or reduced lunches (59%). The sample included 311 typical readers (48.75%) and 578 struggling readers who were receiving either free or reduced price lunch.

Measures and Procedures

ORF. The ORF assessment was designed to measure oral reading of connected text and to monitor reading progress in Grades 6 through 8 for the larger project in which the current study was embedded. A total of 99 graded passages were constructed, of which 35 were used in the present study. These varied in length from 108 to 591 words and in Lexile difficulty levels from 390L to 1050L. The words read correctly per minute score was computed for the first 60 s of reading (1-min fluency). More detailed information about the passages and the assignment of students to passages, including a graphical depiction of the cross-classified nature of the design is available in the supplemental materials.

Correlations across passages at the first occasion of measurement ranged from .73 to .90 for Grade 6 students, from .79 to .91 for Grade 7 students and from .79 to .93 for Grade 8 students, indicating excellent alternate forms reliability. Descriptive statistics for the 35 passages at each wave are provided online in the supplemental materials in the interest of space.

Reader characteristics. Measures of reader characteristics were selected based on their widespread use within the component skills literature and their strong psychometric properties. Only measures collected at the first occasion of measurement are included in this set. That is, we do not include any person characteristics measured as time varying covariates other than time.

Word-level efficiency. The Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Raschote, 1999) comprised sight word and phonemic decoding efficiency subtests designed to measure accuracy and speed of reading real words and regular nonwords, respectively. Alternate form reliability is high for ages 11 to 18 years (viz., .91 to .97).

Decoding. The Test of Silent Word Reading Fluency (TOSWRF; Mather, Hammill, Allen, & Roberts, 2004) measured word identification and speed by having students demarcate word boundaries for 3 min within a continuous sequence of letters that progress in order of reading difficulty. Test–retest reliabilities for age 8 to 12 years exceed .90.

Verbal knowledge. The verbal knowledge subtest of the Kaufman Brief Intelligence Test–second edition (KBIT-2; Kaufman & Kaufman, 2004) assessed receptive vocabulary and general/verbal knowledge. Internal consistency for ages 11 through 18 years ranged from .89 to .94.

Listening comprehension. The listening comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (Williams, 2001) was used to assess understanding of spoken language. Internal consistency for ages 4 through 18 years ranged from .91 to .99.

Text characteristics. Text characteristics for fluency passages were measured using Coh-Metrix and the Lexile Analyzer. The average word frequency, average sentence length, narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion of the ORF passages were measured utilizing the Coh-Metrix Text Easability Assessor (Graesser, McNamara, & Kulikowich, 2011). The average log word frequency measure as opposed to untransformed average word frequency was used because the logarithmic transformation corrects the distribution of word frequency so that it approximates a normal distribution, as well as has a linear fit with reading times (Graesser, McNamara, Louwerse, & Cai, 2004). The Lexile level of the passage was obtained via the Lexile Analyzer which is available from Metametrics. For more detailed descriptions of these measures, see Kulesz et al. (2016).

Statistical Analysis

The main objective of the study was to demonstrate the utility of modern mixed models for integrating the TDF and CSF reading frameworks into a CVR*i* in a developmental context. To do so, we fit a series of cross-classified random effects multilevel models using SAS PROC MIXED/HPMIXED (SAS Institute, 2013) and using the LME4 package in *R* (see Note 3). To estimate this multivariate, longitudinal, cross-classified random effects model, we organized the dataset into a hyper univariate layout, such that each observation on an individual represented a distinct row in the data table that was uniquely identified by the individual, the occasion of measurement, and the passage that was read. This data table consisted of over 23,000 rows reflecting the fact that each of 1,518 students has a minimum of nine observations and a maximum of 23 observations (see Note 4). The modal number of observations is 19 (n = 1,376), although n = 115 cases had a total of 23 observations. In all, a total of 32 sequences of passages were read at any given occasion; of these, 16 consisted of five passages and 16 consisted of three passages. Across all occasions of measurement, 68 combinations of sets of three or five passages were read, 22 in Grade 6, 24 in Grade 7, and 22 in Grade 8, respectively.

We did not attempt to address the clustering of individuals within schools given the relatively small number of schools in the design and the complexity of the models even without consideration of clustering. With that caveat, we approached the modeling in several phases. (a) *Unconditional model*—we estimated a fully unconditional model with random intercepts at the person and passage levels and random residual errors, reflecting the degree to which individual scores are not simply a function of person and passage means. (b) *Text features*—we attempted to determine the improvement in model fit by taking into consideration the text features as a set. Inclusion of text features should reduce the variance in passage intercepts to the extent that text features influence mean performance across passages. Because text features varied across passages and not within passages, the slope for text features cannot vary randomly at the passage level, but can vary randomly at the person level. These random slopes at the person level indicate the degree to which a given text feature affects readers differentially. Thus, in this second phase, we also consider that features such as text-type (expository/narrative) and Lexile difficulty exerted random effects at the student level. (c) *Person characteristics*—in this phase, we incorporated person characteristics into the model. This phase is expected to reduce variability in the person intercepts as well as residual variance to the extent that characteristics of individuals are related to ORF. This model incorporated measures of component skills but also the reader's classification (typical vs. struggling reader) and their grade in school. For text features found to have slopes that varied randomly across students (i.e., text features that exert differential effects on students), we also considered whether student characteristics explained some or all of the variability in slopes with respect to text features. (d) *Developmental phase*—Finally, we incorporated the developmental nature of the design to examine change in ORF over time within student and to examine the possibility that development is heterogeneous across students. In this phase, we also considered the correlation between random slopes for growth and for passage features (Lexile difficulty and expository text).

In Phases 1 to 3, developmental variability in ORF contributed to the residual variance at the person and passage levels, and the within-person error. All models in Phases 1 to 3 assumed a simple, uncorrelated, residual error structure with homogeneous variance. For the final model, we examined the distribution of Level 1 residuals to assess the veracity of model assumptions about residual distributions.

Explanatory variables were selected *a priori* and utilized both reader characteristics, text features, and time to explain variation in reading fluency. Reader characteristics included KBIT-Verbal Knowledge, listening comprehension, decoding, and word-level efficiency measured by the TOWRE. Text features included sentence length, concreteness, cohesion (deep and referential), narrativity, syntactic simplicity, and frequency, Lexile level, and a dichotomous variable indicating whether the passage was narrative or expository. In prior work, we have found that both the dichotomous classification and the continuous measure of narrativity, which applies to both narrative and expository texts, have been predictive of text difficulty.

To estimate the model, we centered time at the second occasion of measurement and measured time in months of instruction during the year. Observations were generally spaced about 6 weeks apart, beginning in Month 2 of the school year and ending in Month 8. Thus, the intercept included in our models with time reflected expected ORF between 3 and 4 months into the academic year. Months of instruction entered the models as a linear effect; we observed no evidence that time effects were curvilinear over the course of one school year. In all models, we used grand mean centering for static covariates at the person and passage level, with the exception of Lexile difficulty, which we centered at 700L. Thus, in models with covariates, the intercept reflected expected performance for a student of average ability as measured by the component skills on a narrative passage that was average in terms of deep and referential cohesion, syntactic simplicity, and so on and measured 700L in difficulty. In addition to centering, we also divided the Lexile score by 100. Thus, a unit change on the Lexile difficulty scale in the models coincided with a change of 100L on the original Lexile scale. From a practical standpoint, these transformations imply that a slope of −6.0 for the Lexile variable in the model implies that fluency declines by six words per minute for an increase in text difficulty of 100L.

Results

Descriptive Statistics

The sampling strategy was intended to yield a higher proportion of struggling readers than typical readers within

Table 1. Reader Characteristics for Typical and Struggling Readers (N = 1,513).

Reader characteristic		Typical			Struggling		
		6th	7th	8th	6th	7th	8th
Sample size	N	230	167	251	350	194	321
Listening comprehension	M	11.44	11.98	12.66	9.41	9.57	10.34
	SD	(2.09)	(2.06)	(2.12)	(2.04)	(2.26)	(2.25)
KBIT-Verbal Knowledge	M	35.99	38.26	41.08	30.31	31.51	34.15
	SD	(5.16)	(4.90)	(5.77)	(6.55)	(6.37)	(6.07)
SWAN inattention	M	−6.18	−5.98	−4.37	2.76	3.06	3.30
	SD	(12.50)	(12.72)	(11.46)	(10.76)	(11.88)	(10.97)
Hyperactivity	M	−6.37	−6.73	−5.03	0.85	0.52	0.62
	SD	(12.51)	(12.37)	(12.19)	(10.66)	(12.07)	(11.55)
Silent reading efficiency	M	99.97	100.28	98.38	85.77	83.42	79.21
	SD	(12.30)	(12.27)	(14.31)	(11.40)	(13.28)	(13.53)
Phonemic decoding efficiency	M	105.24	104.84	106.43	94.40	95.16	93.71
	SD	(14.40)	(13.57)	(12.65)	(14.67)	(16.40)	(15.02)
Sight word efficiency	M	102.93	101.81	101.67	92.69	91.87	91.40
	SD	(12.45)	(10.52)	(11.40)	(11.20)	(11.90)	(11.02)

Note. For modeling purposes, measures were grand mean centered prior to inclusion in models. KBIT = Kaufman Brief Intelligence Test.

each grade. Gender was associated with reading classification such that the proportion of females was higher among typical (56%) as compared with struggling (47%) readers ($\chi^2_{(1)} = 12.7, p < .004$). Race was also associated with reading classification ($\chi^2_{(3)} = 68.5, p < .001$) with the proportion White higher among typical readers (29%) than struggling (14%) readers, and the proportion Hispanic higher among struggling readers (typical: 27%, struggling: 42%).

Table 1 summarizes the reader performance and behavior characteristics. For all characteristics except for reading efficiency, differences between typical and struggling readers were consistent across grades. That is, the interaction between grade and reader group in a two-way analysis of variance (ANOVA) was not significant, $p > .05$. For reading efficiency, differences between reader groups varied across grades (two-way interaction: $p < .001$).

Text characteristics such as sentence length and measures of text difficulty for each of the 35 passages are available in the online supplemental materials. The passages on average had a sentence length of around 11 words with a spread of nearly three words. Means and standard deviations for other text features were as follows: average word concreteness (M = 74.87, SD = 22.35), deep cohesion (M = 67.9, SD = 22.35) and referential cohesion (M = 36.88, SD = 16.1), narrativity (M = 61.25, SD = 25.84), simplicity of syntax (M = 82.57, SD = 14.05), and frequency of words measured on the log scale (M = 3.02, SD = 0.08). Without exception, average fluency performance increased as a function of grade on all passages read by students in more than one grade. In the interest of space, we do not report descriptive statistics for individual passages here but provide them in the supplemental materials that are available online.

*The CVR**i**: Multilevel Models*

Phase 1: Unconditional model. The first phase of multilevel modeling involved estimating an unconditional model with random intercepts at the student and passage level. Random effects for the unconditional model are presented in Table 2, which shows that most of the variability in the data resides at the person level, primarily between individuals. The variance component for persons was over 1,400, while the variance component for passages was 205, and the within-person error variance was 304. Thus, approximately 73% of the variance is between individuals (intra-class correlation [ICC] = .73), 11% is at the passage level, and 16% of the variance is within individuals.

Phase 2: Conditional passage effects. Adding explanatory variables for passages into the model reduces the passage variance from 205.32 to 72.07, or approximately 65%. The residual passage variance increases slightly to 73.18 when some nonsignificant passage features are removed from the model, and increases again to 78.53 when we allow for the possibility that Lexile difficulty and expository text effects vary across individuals. Of some interest in the random effects are the magnitudes of the random slopes for Lexile difficulty (variance = 2.90) and expository text (variance = 56.34), and the correlations among the random effects. For example, we see that the effect of Lexile difficulty is negatively correlated (r = −.66) with the person intercepts as is the effect of expository text (r = −.46), and the two slopes are modestly positively correlated (r = .27). These negative correlations tell us that these text features have smaller slopes for students reading at higher average fluency levels. However, both features negatively impact fluency. Specifically, as the

Table 2. Random Effects for Unconditional and Conditional Models for Passage, Person, and Developmental Effects.

		Student level		Passage level		Error
Model	Source	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*
Unconditional	Intercept	1,405.54	51.63	205.32	49.91	304.29
Passage effects models						
1. Full text model	Intercept	1,406.56	51.60	72.07	20.11	276.45
2. Reduced text model	Intercept	1,406.58	51.60	73.18	19.32	276.45
3. Reduced text/random slopes	Intercept	1,499.98	38.73	78.53	8.55	257.61
	Lexile slope	2.90	1.70			
	Expository slope	56.34	7.51			
Person effects models[a]						
1. Full person model	Intercept	294.05	11.43	71.95	18.89	273.73
Developmental effects models[b]						
1. Fixed growth slopes	Intercept	296.02	11.42	61.09	16.13	233.24
2. Random growth slopes	Intercept	275.99	16.61	57.69	7.60	205.98
	Growth slope	4.88	2.21			

[a]Person models include explanatory variables for passage intercepts from the reduced text model (2). [b]Growth models include explanatory variables for random person and passage intercepts.

Lexile difficulty increases, students read more slowly. Similarly, students read expository text more slowly than narrative text. Consequently, the negative correlations with intercepts for these features actually imply that better readers show greater impacts of the text being expository as compared with narrative and greater impacts of text difficulty as measured by the Lexile level of the text.

The full model for passage effects included effects for Lexile difficulty level, expository text as a dichotomous indicator, narrativity as a continuous indicator, sentence length, word frequency, deep cohesion, referential cohesion, syntactic simplicity, and word concreteness. The reduced text model eliminated effects for sentence length and word frequency, which were highly correlated with the Lexile difficulty level, and effects for deep cohesion, but retained effects for Lexile difficulty, word concreteness, referential cohesion, syntactic simplicity, and narrativity. Although referential cohesion and syntactic simplicity effects were not statistically significant in this model (t = 1.68 and −1.34, respectively, $df > 120$, $p > .05$), they were retained in the model.

Phases 3 and 4: Person effects and developmental effects. The third phase of model fitting incorporated explanatory variables for the person intercepts, that is, individuals' average reading fluency levels estimated at the second measurement occasion. Estimates of random effects for the Phase 3 and 4 models of person effects and of development are also presented in Table 2 (see Note 5). Including all of the component skills in the model reduced the variance in person intercepts to 294.05 from 1,405.54, a reduction of 79.1% (see full person model in Table 2). When a fixed linear slope for instructional months is introduced to account for average linear growth in reading fluency over the five time points, the variance for person adjusts upward slightly to 296.02. Importantly, the residual variance in passage intercepts decreases from 71.95 to 61.09, a reduction of 15%, while the residual error decreases from 273.73 to 233.24, a reduction of 14.8% (see fixed growth slopes model in Table 2). Thus, approximately 15% of the residual variance in passage intercepts and residual error in the text models of Table 2 is due to the failure of those models to account for the average change over time. Allowing the rates of linear growth to vary across students (i.e., the random slopes growth model) reduces the variance in person intercepts to 275.99, which is a reduction of 6.7%, and further reduces the variance in passage intercepts to 57.69, which is a reduction of 5.6%. The largest percent reduction is seen in the within-person error, which drops from 233.24 to 205.98, or 11.7% (see random growth slopes model in Table 2).

Subsequent to allowing the growth slopes to vary across individuals, we considered possible explanatory variables for the random growth slopes. The final person characteristics included several explanatory variables for the growth slopes, specifically, allowing for the slopes to differ across good and poor readers, and allowing slopes to differ across grades. Slopes were found to differ for good and poor readers, but not to vary across grade in this model. The addition of these two explanatory variables reduced the slope variance by less than 1%. As it stands, we have done little to explain the heterogeneity in growth rates across individuals with these two explanatory variables, especially in comparison with the substantial explanatory power of our predictors for explaining variance in person and passage intercepts, and to a lesser extent the within-person error.

Table 3. Random Effects for Final Conditional Model for Passage and Person Effects.

	Student level		Passage level		Error
Source	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*
Intercept	314.94	17.75	58.59	7.66	186.19
Growth slope	5.46	2.34			
Lexile slope	3.61	1.90			
Expository slope	40.04	6.33			

	Correlations among student-level random effects		
	Intercept	Growth	Lexile
Growth	.11		
Lexile	−.39	−.32	
Expository	−.38	.15	−.01

Note. Model includes explanatory variables for random passage and student intercepts and random slopes for growth, Lexile difficulty, and expository text.

Random effects for a final integrated model are presented in Table 3. This integrated model includes random intercepts for persons and passages, as well as random slopes for growth, and for Lexile difficulty level and expository text, along with explanatory variables for all random slopes and intercepts. In this model that allowed for heterogeneous growth over time and heterogeneous effects of text features (Lexile difficulty and expository text type), we found a substantial reduction in the within-person error from 205.98 to 186.19, or 9.6%, for a total error reduction of 38.8% from the unconditional model that simply allowed for random person and passage intercepts. The total variability in the dataset is approximately 1,960.24 (estimated from a model that includes only a grand mean). Thus, the integrated model in Table 3 accounts for approximately 90.5% of the total variance in the set of fluency scores across all students, passages, and time points (1 − 186.19/1,960.24). However, within the accounted for variance is substantial residual variance in average fluency scores at the person (314/1,499 = 21%) and passage levels (58.59/205.32 = 28.5%), as well as unexplained variability in growth and in the effects of Lexile difficulty and text type. Including random effects for growth as well as for Lexile difficulty and text type substantially reduced the correlations between person intercepts and slopes for Lexile difficulty, and slopes for expository text type. These correlations between person intercepts and passage feature slopes decreased from −.66 and −.46 in the reduced text model in Table 2 to −.39 and −.38, respectively, in the final conditional model (see Table 3). Also, the correlation between Lexile slopes and expository text type slopes was reduced from .27 in the text model to −.01 in the final model. The correlation between person slopes and person intercepts was less affected by the inclusion of random effects for the passage features, changing from .16 to .11 when random effects of the two passage features were included in the model.

Fixed effects for the final models from Phases 2, 3, and 4 are presented in Table 4. Terms included in multiple models tended to have similar effects in all models in which they were included. The exceptions were listening comprehension and the interaction of reader type (typical vs. struggling) and months of instruction, which were significant in the person model, but not in the combined model. The difference between good and poor readers was approximately seven words per minute at the second wave and did not vary across grades. Sixth graders read 123.9 and 129.6 words correctly per minute on average, for poor and good readers, respectively. Seventh graders read about 10.7 words per minute faster than sixth graders, and eighth graders read about 29 words per minute faster than sixth graders. Fluency rates improved about three words per minute, per month on average, but growth rates varied substantially across students as evidence by the residual variance for growth slopes in Table 3. Although rates varied randomly at the student level, rates did not differ on average across grades or reader types in the final model. Effects of text type and Lexile difficulty differed between good and poor readers, and effects of Lexile difficulty differed across grades. Fluency rates declined four words per 100 Lexiles of text difficulty for sixth graders and declined about one half word more per 100 Lexiles for older students. This difference suggests that older students adjust their reading rates more as a function of text difficulty and is consistent with the differences in the Lexile effect and the expository text effect between good and poor readers. In both cases, the effects show that good readers reduce fluency more as Lexile difficulty increases and when reading expository as opposed to narrative text. Both word concreteness and narrativity increased reading fluency by about six words per minute on average for a unit change in concreteness or narrativity (about 1.25 standard deviation units).

Checking model assumptions. Finally, we saved residuals from the final model to investigate the reasonableness of the model's assumptions regarding the distribution of errors. Errors appeared to be symmetric about zero, but appeared to be somewhat heavy tailed relative to the normal distribution. In addition, the model specified that the residual within-person errors are distributed with mean zero, homoscedastic variance, and independence across persons and across time within person. We investigated these assumptions by examining the magnitude of correlations among the residuals. When examining the residual correlations, we found that they were not zero, but

Table 4. Fixed Effects Estimates for Final Text, Person, and Combined Models.

	Final text model			Final person model			Final combined model		
	Estimate	*SE*	*t*	*Estimate*	*SE*	*t*	*Estimate*	*SE*	*t*
Intercept	145.36	2.25	64.5						
Grade 6 poor reader				124.49	2.04	60.8	123.9	2.09	59.2
Grade 6 good reader				127.85	2.16	59.3	129.6	2.20	58.8
Grade 7				10.68	1.56	6.9	10.4	1.61	6.5
Grade 8				29.00	1.39	20.8	28.62	1.43	20.0
Time in months				3.28	0.13	26.1	3.26	0.13	24.3
Lexile difficulty	−5.02	1.61	−3.1	−4.85	1.43	−3.4	−4.01	1.46	−2.7
Expository text type	−15.84	0.36	−44.4	−5.00	0.32	−15.8	−4.18	0.60	−7.0
Word concreteness	5.33	2.31	2.3	6.57	2.05	3.2	6.19	2.09	3.0
Referential cohesion	5.86	3.48	1.7	4.32	3.10	1.4	4.51	3.15	1.4
Narrativity	4.69	2.18	2.2	6.11	1.94	3.2	6.11	1.98	3.1
Syntactic simplicity	−6.89	5.14	−1.3	−5.72	4.57	−1.3	−5.26	4.65	−1.1
Silent reading fluency				0.90	0.05	19.8	0.84	0.04	19.1
Verbal knowledge				−0.46	0.21	−2.2	−0.34	0.21	−1.6
Phonemic decoding				0.69	0.05	15.3	0.73	0.04	16.7
Sight word decoding				0.88	0.06	14.6	0.78	0.06	13.4
Listening comp.				0.48	0.23	2.1	0.35	0.23	1.5
Reader type × Time				0.28	0.14	2.0	0.25	0.15	1.6
Reader type × Lexile							−0.89	0.13	−6.6
Reader type × Expository							−1.67	0.64	−2.6
Reader type × Grade 7				−1.41	2.36	−0.6	−1.21	2.30	−0.5
Reader type × Grade 8				−2.02	2.06	−1.0	−1.63	2.00	−0.8
Grade 7 × Time				0.15	0.18	0.9	0.19	0.20	1.0
Grade 8 × Time				−0.13	0.16	−0.8	−0.10	0.17	−0.6
Grade 7 × Lexile							−0.53	0.20	−2.7
Grade 8 × Lexile							−0.57	0.18	−3.2
Grade 7 × Expository							−0.57	0.85	−0.1
Grade 8 × Expository							−0.19	0.73	−0.3

Note. Terms missing in a column were excluded from that model. Interactions between person characteristics and text characteristics were only included in the combined model. In the person model and combined model, the intercept is replaced with specific intercepts for typical and struggling readers in Grade 6. Reader type is the difference between good and poor readers.

were generally substantially smaller than the original correlations across individual measures. Specifically, the residual correlations showed $r < .2$ in absolute value, with many positive and many negative values and many near zero, whereas the correlations among measures showed $r > .85$ and uniformly positive. Thus, the fixed and random effects of the final integrated model substantially capture the correlations among the individual measures. Nevertheless, to assess the extent to which this misspecification of the distribution of errors was influencing decisions regarding model parameters, we attempted to estimate the final model with an unstructured residual covariance matrix. However, we were unable to estimate the fully specified model with random slopes and unstructured covariance matrix. Consequently, we were unable to fully assess the extent to which conclusions from the individual models were impacted by misspecification of the error structure.

Discussion

The SVR has guided the work of reading researchers from the CSF for over 30 years. It is, by far, the single most widely used framework for conceptualizing the process of reading comprehension from the standpoint of the essential skills that readers must use to understand written language. At the same time, the text and discourse reading framework has helped reading researchers to investigate the properties of texts that influence how readers form a mental model of the text and how text features can complicate or facilitate readers' comprehension. For the most part, these frameworks have guided reading research in isolation with only a handful of exceptions, including some recently that have made use of advances in statistical modeling (Kulesz et al., 2016). In the present study, we have shown how the SVR, or more broadly the CSF, can be extended and integrated with the TDF to form a CVR*i*. As such, the CVR*i* is both a component skills and text features model, and most importantly, the CVR*i* is a personalized model of reading, capable of addressing heterogeneity in the effects of text features, in the development of reading comprehension, the effects of motivation, and variation in the demands on the reader as a result of the specific purpose for reading a given text. Our empirical example showed how the modeling of fluency as a proxy for comprehension provides evidence for the fact that readers do not deploy their cognitive resources in homogeneous ways to solve the reading task. Specifically, we found evidence of heterogeneous development of fluency across children in sixth through eighth grades, but also found evidence of heterogeneous effects of text features, such as the Lexile difficulty level and the type of text. We did not conduct an exhaustive search of text features or person features to ascertain all such features whose effects vary across readers. Indeed, we would say that we have only begun to scratch the surface in this regard and much additional work needs to be done.

The fact that text features may affect readers differently, and/or that readers may deploy their cognitive abilities differently from one another in comprehending text has important implications for intervention research. In the SVR and in all CSF models, the presumption exists that what distinguishes good from poor readers is their relative standing on the component skills. Improve an individual's standing on the component skills and their reading will improve commensurately! However, if readers deploy their component skills in different ways, or are differentially impacted by the effects of text features, then simply changing the reader's standing on the component skills may not change comprehension to the degree predicted by the between person regression coefficient that relates the component skill to comprehension in component skills research. Recognition of this fundamental distinction between within person and between person covariation lies at the heart of personalized medicine, and by extension at the heart of personalized education and educational interventions. We should add that personalized education and medicine are modern day extensions of earlier interaction models in psychology and education, such as aptitude/treatment and attribute/treatment interaction models (Cronbach, 1957; Cronbach & Snow, 1977; see Note 6). In the present case, we are distinguishing the models proposed here from notions of learning styles (Pashler, McDaniel, Rohrer, & Bjork, 2008); rather, what we are proposing are performance styles, or how individuals deploy their cognitive skills to address the motivational, cognitive, and linguistic challenges posed in text, and the implications that such differences might have for working with students to improve their comprehension. That is not to say that all intervention or all learning is idiographic, but merely an acknowledgment of the possibility that the functions that relate component skills and text features to comprehension may comprise both aggregate and individualized elements. In fact, the failure to find consistent evidence of learning styles (Pashler et al., 2008) might stem, in part, from the involvement of aggregate (i.e., common across all individuals) elements in learning and performance functions that lowers power for identification of the individualized elements. Only with intensive individual level data collections on large samples of subjects can the two sources be isolated.

The CVR*i* is consistent with at least two of the National Science Foundation's Big Ideas for the 21st century, specifically Understanding the Rules of Life: Predicting the Phenotype, and Harnessing Data for 21st Century Science and Engineering. Through the use of personalized models and developmental functions, the CVR*i* offers a richer description of the reading phenotype than can be readily captured by the SVR or any components skills framework that ignores the contribution of text features to comprehension and the heterogeneity that exists across individuals in the parameters of the model/function. Exploiting these personalized models requires massive datasets with substantial variation at the individual level. That is, we require datasets with many replicate observations on each individual, preferably observations that are coded with extensive information about the task and the stimulus. As large educational datasets become more ubiquitous and are increasingly publicly available, and/or as datasets are designed or assembled with models like the CVR*i* in mind, more complete descriptions of the reading phenotype become possible. In some cases, these datasets may be linked to neuroimaging and/or genetic or epigenetic data, or other developmental information, making richer, dynamic descriptions of the reading phenotype increasingly possible.

Our approach to the modeling was limited by the assumptions regarding the distribution of residuals. Although we were able to substantially reduce the correlations among residuals, we were not able to eliminate them entirely, and estimating the models with random slopes and non-diagonal, nonhomogeneous error distributions proved to be computationally problematic. We expect this limitation to be overcome in future work, both through improved computational approaches and also through the inclusion of other person and text features, such as text familiarity, that might reduce the correlations among residuals even further, making the assumptions of homogeneous, diagonal errors more plausible.

Finally, our study involved the use of reading fluency as a proxy for comprehension, but the CVR*i* is not restricted to ORF. Indeed, the model could be used with direct measures of reading comprehension, measured either continuously or categorically, such as in the case of explanatory item-response models for reading comprehension test items (Kulesz et al., 2016). Unfortunately, we did not have at our disposal a novel dataset with many replicate observations of reading comprehension where we had also coded the instruments for their text features. As reading research advances into the 21st century, we expect that researchers will not be content to study variation in reading processes using models that focus exclusively on variability between individuals. Rather, we expect that reading research will continue to evolve toward models that are capable of integrating variation in readers and texts and that capture how individual readers deploy their cognitive resources to meet the demands of the reading task as well as those characteristics of the reading process that are consistent across readers and contexts.

Acknowledgments

The authors would also like to thank our colleagues, Lee Branum-Martin and Elena Grigorenko, and the journal reviewers for their helpful comments, and accept full responsibility for any misinterpretations. This article includes supplemental materials, which can be accessed at https://www2.times.uh.edu/files/FrancisKuleszBenoit_CVRi_RSE_Supplement.pdf.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by grants P50 HD052117, Texas Center for Learning Disabilities, from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and R305F100013, Reading for Understanding: Understanding Malleable Cognitive Processes and Integrated Comprehension Interventions for Grades 7 to 12, funded under the Reading for Understanding Research Initiative by the Institute for Education Sciences of the U.S. Department of Education. The content and opinions expressed herein are the responsibility of the authors and do not represent views of the Institute or the U.S. Department of Education nor the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health.

Notes

1. The authors would like to thank Dr. Lee Branum-Martin, Georgia State University, for this helpful insight.
2. For time varying covariates, it is also often useful to code these into a static component, such as the average value or the starting value for the individual, and the deviations from that value over the replicate observations. In our formulation, the static component would be placed in X and the deviations, which vary over time, in D_t. By coding time varying covariates in this way, we can isolate the contribution of the covariate to interindividual differences, which is captured by the component placed in X, and the contribution of the covariate to intraindividual differences, which is captured by the component placed in D_t.
3. Additional details on the use of MIXED, HPMIXED, and LME4 (Bates, Maechler, & Dai, 2008) is available from the authors.
4. We dropped one observation associated with a fluency score of 1,000 at one measurement occasion. All remaining observations for the individual student with the anomalous score were retained in the analysis.
5. NB: All models for person effects in Table 2 included the fixed effects for the explanatory variables for passage intercepts from the reduced text model (2). Developmental models in Table 2 include the explanatory variables for random person and passage intercepts included in the full person model. Random slopes for Lexile difficulty and expository text are excluded from the person models and developmental models in Table 2.
6. The authors would like to thank Elena Grigorenko for her insights on the broader context in which the CVR*i* is situated.

ORCID iD

David J. Francis https://orcid.org/0000-0003-3944-3274

References

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M. A., & Kulesz, P. (2016). Validation of the direct and mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology*, *44*, 68–82.

Barnes, M. A., Dennis, M., & Haefele-Kalvaitis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inference in children from six to fifteen years of age. *Journal of Experimental Child Psychology*, *61*, 216–241.

Barth, A. E., Tolar, T. D., Fletcher, J. M., & Francis, D. (2014). The effects of student and text characteristics on the oral reading fluency of middle-grade students. *Journal of Educational Psychology*, *106*, 162–180. doi:10.1037/a0033826

Bates, D., Maechler, M., & Dai, B. (2008). The lme4 package, version 0.999375-26 [Software]. Retrieved from https://cran.r-project.org/web/packages/lme4/

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*, *29*, 137–164.

Boscardin, C., Muthen, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology*, *100*, 192–208.

Cain, K., & Oakhill, J. (2007). Reading comprehension difficulties: Correlates, causes, and consequences. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41–75). New York, NY: Guildford Press.

Cain, K., & Oakhill, J. (2009). Reading comprehension development from 8 to 14 years: The contribution of component skills and processes. In R. K. Wagner, C. Schatschneider & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp. 143–176). New York, NY: Guildford Press.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*, 31–42.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 25–40). Mahwah, NJ: Lawrence Erlbaum.

Cronbach, L. J. (1957). Two disciplines of scientific psychology. *American Psychologist*, *12*, 671–684.

Cronbach, L. J., & Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.

Francis, D. J., Barr, C. D., Benoit, J. S., & McIntyre, T. M. (2017). Issues in research methodology on educator stress. In T. M. McIntyre, S. E. McIntyre & D. J. Francis (Eds.), *Educator stress: An occupational health perspective* (pp. 439–470). New York, NY: Springer.

Francis, D. J., Fletcher, J. M., Catts, H., & Tomblin, B. (2005). Dimensions affecting the assessment of reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369–394). Mahwah, NJ: Lawrence Erlbaum.

Garcia, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*, 74–111.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency based indicators of foundational reading skills for third-grade high stakes outcomes. *Scientific Studies of Reading*, *5*, 257–288.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In J. V. Oakhill & C. Cornoldi (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1–13). Mahwah, NJ: Lawrence Erlbaum.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6–10. doi:10.1177/074193258600700104

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*, 371–398.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analysis of text characteristics. *Educational Researcher*, *40*, 223–234.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*, 193–202.

Hannon, B., & Daneman, M. (2009). Age-related changes in reading comprehension: An individual-differences perspective. *Experimental Aging Research*, *35*, 432–456.

Joshi, M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, *21*, 85–97.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test Manual* (2nd ed.). San Antonio TX: Pearson.

Kulesz, P. A., Francis, D. J., Barnes, M. A., & Fletcher, J. M. (2016). The influence of properties of the test and their interactions with reader characteristics on reading comprehension: An explanatory item response study. *Journal of Educational Psychology*, *108*, 1078–1097. doi:10.1037/edu0000126

Lubke, G. H., & Muthen, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *101*, 21–39.

Lynch, J., & van den Broek, P. (2007). Understanding the glue of narrative structure: Children's on- and off-line inferences about characters' goals. *Cognitive Development*, *22*, 323–330.

Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2004). *Test of Silent Word Reading Fluency (TOSWRF)*. Austin, TX: Pro Education.

McNamara, D. S. (2001). Reading both high and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*, 51–62.

McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty: Across the ages and genres. In J. P. Sabatini, E. Albro & R. T. O'Reilly (Eds.), *Measuring up* (pp. 89–119). Lanham, MA: Rowman & Littlefield Education.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–288.

Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1–4. *Scientific Studies of Reading*, *9*, 85–116.

Meyer, B. J. F., & Ray, M. N. (2011). Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education*, *4*, 127–152.

Millis, K., Graesser, A., & Haberlandt, K. (1993). The impact of connectives on the memory for expository prose. *Applied Cognitive Psychology*, *7*, 317–339.

Nation, K., & Snowling, M. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, *27*, 342–356.

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*, 91–121.

Oliviera, D. G., da Silva, P. B., Dias, N. M., Seabra, A. G., & Macedo, E. C. (2014). Reading component skills in dyslexia: Word recognition, comprehension, and processing speed. *Frontiers in Psychology*, *5*, Article 1339.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*, 228–242.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, *9*, 105–119.

Perfetti, C. A., & Adolf, S. M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In

J. P. Sabatini, E. Albro & R. T. O'Reilly (Eds.), *Measuring up* (pp. 3–20). Lanham, MA: Rowman & Littlefield Education.

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*, 22–37.

Pyle, N., Vasquez, A. C., Lignugaris/Kraft, B., Gillam, S., Reutzel, D. R., Olszewski, A., & Pyle, D. (2017). Effects of expository text structure interventions on comprehension: A meta- analysis. *Reading Research Quarterly*, *52*, 469–501.

Roberts, G., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly*, *20*, 304–317.

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Not just speed reading: Accuracy of the DIBELS oral reading fluency measure for predicting high-stakes third grade reading comprehension outcomes. *Journal of School Psychology*, *46*, 343–366.

SAS Institute. (2013). *SAS 9.4 guide to software updates*. Cary, NC: Author.

Scammacca, N. L., Roberts, G. R., Vaughn, S. R., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities*, *48*, 369–390.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265–282.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: RAND Corporation.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947.

Tighe, E. L., & Schatschneider, C. (2014). A dominance analysis approach to determining predictor importance in third, seventh, and tenth grade reading comprehension skills. *Reading and Writing*, *27*, 101–127.

Torgesen, J. K., Wagner, R., & Raschote, C. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed.

van den Broek, P. (2012). Individual and developmental differences in reading comprehension: Assessing cognitive processes and outcomes. In J. P. Sabatini, E. Albro & R. T. O'Reilly (Eds.), *Measuring up* (pp. 39–58). Lanham, MA: Rowman & Littlefield Education.

van den Broek, P., Lorch, E. P., & Thurlow, R. (1996). Children's and adult's memory for television stories: The role of causal factors, story-grammar categories and hierarchical level. *Child Development*, *67*, 3010–3029.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skill model of reading development. *Scientific Studies of Reading*, *11*, 3–32.

Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*, 407–423.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology*, *30*, 73–87.

Williams, K. T. (2001). *Group reading and diagnostic evaluation*. New York, NY: Pearson.

Yap, M. J., Tse, C. S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency revealed by RT distributional analyses: The role of lexical integrity. *Journal of Memory and Language*, *61*, 303–325.