# Decoding and reading comprehension: A test of the Decoding Threshold Hypothesis

Zuowei Wang, John Sabatini, Tenaha O'Reilly, Jonathan Weeks

Zuowei Wang (corresponding author), Educational Testing Service, 660 Rosedale Road MS 13-E, Princeton, NJ 08541, USA. Phone: 609 252 8375 email: zwang@ets.org

## Acknowledgements

The research reported here was supported by the Institute of Education Sciences (IES), U.S. Department of Education, through Grant R305F100005 to the Educational Testing Service (ETS) as part of the Reading for Understanding Research (RFU) Initiative, and Grant R305G040065. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We would like to thank Don Powers, Hyo Jeong Shin and Laura Halderman for providing helpful feedback on earlier versions of this manuscript.

© 2018, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article to be published in the Journal of Educational Psychology. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/edu0000302

Suggested citation:

Wang, Z., Sabatini, J., O'Reilly, T., Weeks, J. (in press). Relation between decoding and reading comprehension: a test of the decoding threshold hypothesis. *Journal of Educational Psychology.*

Abstract

We report results of two studies examining the relation between decoding and reading comprehension. Based on our analysis of prominent reading theories such as the Simple View of Reading (Gough & Tunmer, 1986), the Lexical Quality Hypothesis (Perfetti & Hart, 2002) and the Self-teaching Hypothesis (Share, 1995), we propose the Decoding Threshold Hypothesis, which posits that the relation between decoding and reading comprehension can only be reliably observed above a certain decoding threshold. In Study 1, the Decoding Threshold Hypothesis was tested in a sample of over 10,000 Grade 5-10 students. Using quantile regression, classification analysis (Receiver Operating Characteristics) and broken-line regression, we found a reliable decoding threshold value below which there was no relation between decoding and reading comprehension, and above which the two measures showed a positive linear relation. Study 2 is a longitudinal analysis of over 30,000 students' reading comprehension growth as a function of their initial decoding status. Results showed that scoring below the decoding threshold was associated with stagnant growth in reading comprehension. We argue that the Decoding Threshold Hypothesis has the potential to explain differences in the prominent reading theories in terms of the role of decoding in reading comprehension in students at Grade 5 and above. Furthermore, the identification of decoding threshold also has implications for reading practice.

Key words: decoding threshold; self-teaching; simple view of reading; lexical quality hypothesis; reading development

## Educational Impact and Implications Statement

This study supported the Decoding Threshold Hypothesis, which posits that the relation between decoding and reading comprehension becomes unpredictable when decoding falls below a threshold. As many as 38% of Grade 5 students and 19% of Grade 10 students in our sample were below the decoding threshold. These students did not make any progress in their reading comprehension score in the following three years; their peers did. Thus, the decoding threshold provides a way to identify students whose reading comprehension will likely remain poor unless their decoding can be improved to a level above the decoding threshold.

Decoding and reading comprehension: A test of the Decoding Threshold Hypothesis

The ability to decode printed texts into meaningful language units is critical to reading. This assertion is supported by a number of prominent reading theories, including the Simple View of Reading (Gough & Tunmer, 1986), the Lexical Quality Hypothesis (Perfetti & Hart, 2002), and the Self-Teaching Hypothesis (Share, 1995). The authors of the Simple View of Reading (SVR) hypothesized that decoding and language comprehension are the two primary components of reading comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990). The Lexical Quality Hypothesis (LQH) posits that successful comprehension is dependent on "accessible, well specified and flexible knowledge of word forms and meanings" (Perfetti & Adlof, 2012, p.9), and decoding provides the mechanism for acquiring this lexical knowledge. Similarly, the Self-Teaching Hypothesis (STH) proposes that decoding allows developing readers to transform unfamiliar printed letter strings into sounds that they can recognize from their spoken language, and this ongoing decoding process provides opportunities for the reader to internalize the orthographic features of new words, a key process in learning to read (Share, 1995).

While the authors of these influential reading theories agree on the critical role of decoding in reading, they differ in predicting the exact relation between decoding and reading comprehension. For the SVR, because reading comprehension is the product of decoding and language comprehension, it is expected that decoding predicts reading comprehension in a similar way throughout the spectrum of language comprehension ability. In other words, higher decoding ability is generally associated with better reading comprehension. For the LQH, low ability in decoding consumes the cognitive resources for higher level processing and thus results in limited reading comprehension. Therefore the LQH seems to imply that the relation between decoding and reading comprehension might change at some point along the continuum of decoding proficiency. For the STH, decoding provides opportunities for the further development of reading comprehension. Thus, decoding ability predicts the development of reading comprehension. Collectively, these predictions are not necessarily mutually exclusive, but rather reflect different aspects of the relation between decoding and reading comprehension. To further explore these theoretical aspects in older students, the current research takes advantage of a large sample to examine this relation with both cross-sectional (over 10,000 students) and longitudinal data (over 30,000 students and 50,000 observations).

We believe having a more comprehensive understanding of the role decoding plays in reading and reading development will make contributions to both reading theories and practice. In the following sections, we first provide brief reviews of SVR, LQH and STH, based on which we propose a Decoding Threshold Hypothesis. To test this hypothesis, we use a multifaceted decoding measure, and examine whether the relation between decoding and reading comprehension is linear at a single point in development, and if not, whether we can find a reliable, lower threshold[1] point where the relation changes. Finally, to examine the validity of the decoding threshold, we then track the reading development of students whose initial decoding is above or below the decoding threshold over a period of three years (up to four time points).

**Decoding in the Simple View of Reading (SVR)**

Gough and Tunmer (1986) first proposed the SVR in an effort to address conceptual confusions in studying the role of decoding in reading. According to the authors, reading

---

[1] In this study, we use the term threshold to represent a lower bound inflection point where the nature of the linearity of the relationship between decoding and comprehension changes. Hypothetically, an upper bound inflection point is also plausible – a point beyond which decoding is at mastery or ceiling levels, and consequently its relationship to comprehension changes again. This upper bound threshold hypothesis is not explored here.

comprehension (R) is the product of decoding (D) and linguistic comprehension (C), $R = D \times C$. This equation has several implications: 1) both decoding and linguistic comprehension are necessary but not sufficient conditions of reading comprehension; 2) decoding and linguistic comprehension are two separable constructs —otherwise one can propose an even simpler view predicting reading comprehension by either construct; and 3) reading comprehension can be reliably predicted given one's decoding ability and comprehension ability.

In the original instantiation of the SVR, decoding is defined as the "knowledge of the spelling-sound correspondence rules of English", which can be measured by having subjects pronounce pseudowords (e.g. clard). Gough and Tunmer (1986) acknowledged that the ability to read orthographically regular pseudowords does not always predict word recognition, because of the fact that sight to sound correspondences in English are not regular and consistent (e.g., *though*, *thought* both start with 'thou', but result in different phonological representations of the 'ou' sound when pronounced in real words), but they argued that knowledge of letter-sound correspondence rules is necessary for the reader to recognize the majority of words. This view was extended in the Self-teaching Hypothesis (Share, 1995), which argued that decoding is actually necessary for the reader to learn all words, orthographically regular or not.

Since it was proposed, the basic premises of the SVR have received strong support from empirical studies (Catts & Weismer, 2006; Chen & Vellutino, 1997; Hoover & Gough, 1990; Johnston & Kirby, 2006; Sabatini, Sawaki, Shore, & Scarborough, 2010; Savage, 2006; Tilstra, McMaster, Van den Broek, Kendeou, & Rapp, 2009). Specifically, these studies have shown that decoding and linguistic comprehension are sufficient to explain a large portion of variation in reading comprehension.

However, the results are more ambiguous as researchers have explored a range of constructs and measures that constitute decoding and linguistic comprehension, across a wider range of ages and grades. For example, measures of reading fluency or vocabulary have been shown to explain additional, unique variance to reading comprehension beyond decoding and language comprehension, leading some researchers to conclude that they should be considered as subconstructs of SVR (Kirby & Savage, 2008; Sabatini et al., 2010). Other studies have been published where SVR models did not fit the empirical data well. For example, in a study of students from Grade 4, 7 and 9, Tilstra et al. (2009) found that verbal proficiency (measured with a vocabulary definition task) and reading fluency significantly predicted reading comprehension after decoding and linguistic comprehension were controlled. Similarly, Savage (2006) examined the SVR in a group of 15-year-old students who had severe reading delays. He found that verbal ability provided better prediction to reading comprehension than linguistic comprehension if decoding was measured by text reading accuracy. In that study, verbal ability was measured by Word Definitions and Verbal Similarities test (Elliott, Smith, & McCulloch, 1996), in which students needed to verbally define words and explain the similarities of words provided in lists. Additionally, Johnston and Kirby (2006) found that naming speed explained a significant amount of unique variance in predicting reading comprehension beyond decoding and linguistic comprehension in poor readers from Grade 3-5, leading them to argue that the SVR was too simple a model to describe poor readers.

Given the complexities in these empirical findings, several factors seem to be helpful when considering the effective range of application of the SVR model. The first factor is how the SVR constructs were conceptualized, operationalized, and measured across age groups (Kirby & Savage, 2008; Savage, 2006). With student samples from upper elementary and beyond, for example, simple tasks such as one- or two-syllable word decoding may not yield sufficient variation, even among relatively poor readers. The lack of variation may oversimplify the decoding construct. The second factor is reading skills. The SVR model did not fit the data as well in samples comprised mostly of students who had poor reading skills (Johnston & Kirby, 2006; Savage, 2006).

The fact that the SVR does not fit empirical data well at the lower end of reading skills implies that the equation $R = D \times C$ may require some modification when decoding or comprehension is low. This suggested to us that there might be a decoding threshold below which the relation between reading comprehension and decoding does not follow the SVR equation. There is some preliminary evidence suggesting this nonlinear relationship under the SVR framework. According to the statistical modeling work by Chen and Vellutino (1997), which was validated by empirical data, the equation $R = D + C + D \times C$ fitted data better than $R = D \times C$. This casts doubt upon the linear relation between decoding and reading comprehension throughout the ability distribution of readers, as the original SVR equation implies.

**Decoding in the Lexical Quality Hypothesis (LQH)**

Another reading model similar to the SVR was proposed by Perfetti and colleagues (Perfetti & Hart, 2002). This model is more process-oriented, and it provides detailed descriptions about the relations among different reading components. Perfetti’s componential model is similar to the SVR in that both models proposed that reading contains two components: in SVR, decoding and linguistic comprehension, and in Perfetti’s model, the orthographic system and the linguistic system. The two models differ in that whereas the two components in the SVR are treated as if they are independent constructs, the orthographic system and the linguistic system in Perfetti’s model are explicitly interactive, with both word identification and higher level processes of reading (e.g. inference making, comprehension monitoring and strategy use etc.) sharing the same cognitive resources during the process of text comprehension.

The interactive process between word identification and higher level reading processes is described in the LQH (Perfetti & Hart, 2002). According to LQH, the quality of lexical representations plays a critical role in reading comprehension, because successful comprehension is dependent on "accessible, well specified and flexible knowledge of word forms and meanings" (Perfetti & Adlof, 2012, p.9). Thus, word identification not only includes the ability to recognize a word on its surface level, but also the activation of knowledge related to the form and meaning of it. Studies have found that poor readers who appear to have normal decoding and word recognition performance showed differences in the cognitive processing of words, as reflected by experiments using semantic priming and eye-tracking tasks (Nation & Snowling, 1999; Veldre & Andrews, 2014). These results indicate that poor readers need extra processing in word identification due to their low quality of lexical representation. Because the cognitive resources (e.g., attention, executive function, and working memory) for comprehension are limited, ineffective word identification consumes the cognitive resources that would otherwise be available for higher-level processing, such as inference making and comprehension monitoring, and this negatively impacts reading comprehension (Walczyk, Marsiglia, Johns, & Bryan, 2004).

Although Perfetti's model does not specifically include decoding as a component, decoding is part of the word identification system. Thus, in Perfetti's model, we see the relation between decoding and reading comprehension as more complex than that predicted by the SVR. Because efficient decoding and word identification makes room for higher level reading processes, it is expected that problems in decoding will result in problems in higher level processes, thus limiting reading comprehension performance. At the same time, efficient decoding only makes possible, but does not guarantee the operation of higher level processes. Thus, the model suggests that there might be a minimum level of decoding skill before higher level processing is operational, a necessary condition for successful comprehension. Below this decoding threshold, reading comprehension remains limited and there is no obvious relation between decoding and reading comprehension; above this threshold, the relation between decoding and reading comprehension may follow the SVR. This is the Decoding Threshold Hypothesis that we will test in the current studies.

**Role of decoding in the Self-Teaching Hypothesis (STH)**

The STH treats decoding as a central driving factor for reading acquisition (Share, 1995). The STH was proposed to account for the vast number of unfamiliar words the developing reader encounters when learning to read. Nagy and Herman (1987) estimated that an average fifth grader encountered about 10,000 new words in a year. Based on this, Share (1995) argued that the only feasible mechanism that students learn these words is through a mechanism that allows self-teaching. Share further proposed that the only possible self-teaching mechanism is through phonological recoding, or decoding. Successful decoding allows students to “translate” unfamiliar printed words into spoken language, which can be recognized and then learned.

The STH was supported by a number of empirical studies. Cunningham, Perry, Stanovich, and Share (2002) found that successful decoding of novel words predicted orthographic learning in second graders when they read short expository texts. Interestingly, after decoding rate was controlled, neither rapid automated naming nor general cognitive ability remained a significant predictor to orthographic learning. This supported the centrality of decoding in orthographic learning. Studies have also revealed that the self-teaching process happens very quickly. For third graders, a single encounter with a novel orthographic string that was presented in short texts resulted in reliable recall of the orthographic features and the acquired orthographic feature was still remembered after a month (Share, 2004).

The STH is also supported by evidence from children who had decoding difficulties. Juel (1988) followed elementary school students' reading development from Grade 1 through 4. She found that students' poor decoding ability at Grade 1 predicted slow progress in both listening comprehension and reading comprehension. Poor decoders also read less in and outside of school and they expressed less interest in reading. These results reveal a vicious cycle that shows how self-teaching can fail as a result of decoding problems: because of the difficulties in decoding, poor decoders read less; because they read less, their decoding and comprehension skills have fewer opportunities for development compared to their peers.

An important question about the findings in Juel (1988) is at what level decoding ability becomes insufficient to drive the self-teaching mechanism. Share and Shalev (2004) found that both poor and good readers demonstrated orthographic learning that was consistent with predictions of the STH. However, good readers, who had stronger decoding skills, showed more word learning in the process. It follows that if decoding falls below a threshold, word learning would become too slow for normal reading development. Indeed, simulation studies that followed a connectionist model of word learning demonstrated that large differences in learning gains occurred in simulated systems that differed in learning capacity (number of "hidden units"), and the differences were especially salient at early exposures to the learning stimuli (Seidenberg & McClelland, 1989). This implies that some disadvantaged learners, either due to slower learning rate or insufficient exposure to learning materials, will likely experience tremendous challenges to keep up with the word learning of their peers through self-teaching.

The question then is: what is the decoding level below which the self-teaching mechanism becomes so difficult to operate that it becomes essentially non-existent? The exploration of this question has significant implications for the identification of decoding problems that contribute to reading difficulties, especially in older struggling readers. If one can identify insufficient decoding when it hinders the self-teaching mechanism, then proper intervention focused on decoding could help bring affected students back on track. The goal of the current study is to explore whether we can identify such a threshold in decoding in older students (i.e., 5$^{th}$ grade and above).

To summarize, from the three prominent reading theories reviewed in this paper on the role of decoding in reading comprehension, we can conclude the following: the SVR implicitly assumes that decoding is linearly related to reading comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990), but results from a number of recent studies suggest the SVR may need some modification, especially at the lower ability range (Johnston & Kirby, 2006; Savage, 2006; Tilstra et al., 2009). The LQH suggests that a minimum amount of decoding needs to be reached before higher-level reading processes can operate for successful reading comprehension, and thus, below this threshold the relation between decoding and reading comprehension is probably unpredictable (Perfetti & Hart, 2002). Finally, the STH also indicates a decoding threshold for the developing reader to reach for normal reading development (Share, 1995). Collectively, evidence from all these three perspectives calls for a test of the Decoding Threshold Hypothesis.

A necessary (but not sufficient) condition for this Decoding Threshold Hypothesis to hold is that the relation between decoding and reading is not linear. We suggest that this non-linear relation was alluded to in prior studies.

**Evidence suggesting non-linear relation between decoding and reading comprehension**

Keenan, Betjemann, and Olson (2008) found that the relation between decoding and comprehension varied when reading comprehension was measured by different reading tests, and they also found that both chronological age and reading age interacted with decoding when predicting reading comprehension performance. Since the relation between decoding and reading comprehension is dependent on age and reading development, it cannot be linear, at least in a cross-sectional sample.

Evidence from longitudinal studies of reading provides further evidence for a non-linear relation between decoding and reading comprehension. In a longitudinal design, Aarnoutse, Van Leeuwe, Voeten, and Oud (2001) tracked the developmental patterns of students' reading skills from Grade 1 through Grade 6. They found that decoding and reading comprehension showed different patterns of growth. Specifically, the growth rate in decoding showed steady decline through the six years, whereas the growth of reading comprehension showed a peak between Grade 2 and 3. Interestingly, Foorman, Petscher, and Herrera (2018) found that the unique variance of reading comprehension that was explained by decoding also showed a peak around Grade 2 and 3. This asynchronization of growth in decoding and reading comprehension indicates that the relation between decoding and reading comprehension may be non-linear across development.

Other studies have identified non-linear relations between decoding and reading fluency, a measure that has been widely used as a proxy for reading comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Hosp and Fuchs (2005) found that the relation between decoding and reading fluency was weaker in Grade 1 students than that in Grade 2 or Grade 3 students. Similarly, Catts, Petscher, Schatschneider, Bridges, and Mendoza (2008) followed three groups of young children (kindergarten, 1$^{st}$ and 2$^{nd}$ grade) for one academic year and evaluated how their decoding was related to oral reading fluency. They found that the relation between decoding and oral reading fluency was constrained in lower grade students who showed floor performance in decoding. As they grew older the relation between decoding and oral reading fluency became stronger, and the relation was also stronger in children who had higher decoding scores. Given the close relation between oral reading fluency and reading comprehension, these results imply a non-linear relation between decoding and reading comprehension across development.

In short, both theoretical analysis and results from several empirical studies necessitate a systematic investigation of the Decoding Threshold Hypothesis. In the next section, we review prior research to examine measurement issues related to decoding and reading comprehension to set up the context for the current studies.

## Measurement considerations of decoding and reading comprehension

García and Cain (2014) provided a comprehensive literature review examining the moderators of the relation between decoding and reading comprehension. After performing a meta-analysis of 110 studies, they found that the overall corrected correlation between decoding and reading comprehension was .74. Pertinent to this paper is their subsequent analysis on moderating factors, including both reader and assessment characteristics. Here we summarize the key findings from García and Cain (2014) and explain the implications for the current research.

**Readers' age.** Among reader characteristics, García and Cain (2014) found that age was the strongest moderator, with the relation much stronger in younger readers. The corrected correlation between decoding and reading comprehension was estimated at .80 for readers not older than 10 years old, whereas the correlation for older readers was estimated to be .47. The division of age groups in García and Cain (2014) at 10 years old roughly coincides to when decoding instruction in English ceases in the United States (i.e., fourth grade) and many other countries (Chall & Jacobs, 2003). Indeed, teachers often believe that by fourth grade, students have made the transition from “learning to read” to “reading to learn” (Houck & Ross, 2012). This seemingly reasonable belief will likely have consequences for decoding instruction. Teachers who hold this popular belief will naturally shift their instruction away from decoding instruction after Grade 4. One the one hand, this saves instructional time for more advanced reading skills; on the other hand, under this instructional policy, students who still have poor decoding beyond this age will struggle with reading due to their poor decoding skills, making it very difficult for them to catch up with their peers. If there is a significant prevalence of children with poor decoding beyond the elementary grades, then such a policy exacerbates the risk of inadequate reading development for students as they enter the middle grades (Sabatini, Wang, O’Reilly, in press). In the current study, we focus on this older group of students to test this Decoding Threshold Hypothesis. This would not only help identify the group of students whose decoding is still very poor beyond Grade 4, but potentially suggest different paths of instruction. The identification of decoding threshold in this age group might signal the need for decoding instruction for students who might otherwise be misclassified as having problems with only comprehension processes.

**Decoding measures.** With respect to assessment characteristics, García and Cain (2014) found that the way that decoding was measured moderated the strength of relation between the constructs of decoding and comprehension. Specifically, they found that the correlation to reading comprehenion was stronger when decoding was measured by accuracy rather than by speed. Further, the correlation was strongest when decoding was measured by real word reading, followed by pseudo word reading, which in turn was followed by lexical decision (i.e., deciding whether a letter string is a real word or made up word), though all have been used in studies as valid indicators of decoding ability. Importantly, García and Cain (2014) identified significant interactions between age and decoding measures. For example, some decoding measures are more sensitive to the individual differences for specific samples for different age/ability groups. The implication of these findings for the current studies is that if we are to examine the relation between decoding and reading comprehension in a wide range of age/ability groups, an inclusive measure that requires decoding at different grain sizes (i.e., focuses on accuracy of real word recognition and phonological recoding of pseudowords) is more likely to be sensitive to individual differences than a narrow measure. Furthermore, to analyze student change longitudinally, the decoding measure needs to produce scores that can be compared across the age/ability groups (i.e., measure that is verticially scaled). For example, we should be able to compare a high-ability, younger student's score to an average-ability, older student's score.

The decoding measure used in the current study satisfies these requirements. We used the decoding subtest from the RISE battery, an assessment developed by ETS (http://rise.serpmedia.org/index.html). It is a computer-administered reading assessment for Grade 5-10. Each of the subtest forms of the RISE battery have been vertically scaled, using an item response theory (IRT) framework. Thus, scores from students who take different RISE forms (within each of the subtest constructs) are on the same scale, and thus comparable longitudinally, and independent of age/grade level. Because of this vertical scale, RISE subtests are sensitive to the development of students' component reading skills across grade and age level (Sabatini, Bruce, Steinberg, & Weeks, 2015). It has been used in several large scale intervention studies to evaluate the effectiveness of intervention programs, demonstrating its practical sensitivity to instructional gains/changes in student growth (e.g. Kim et al., 2017).

In RISE, decoding is measured in a task design that incorporates real word identification, phonological decoding of pseudo-words, and matching of pseudo-homophones (e.g., whissle) to known words in one's mental listening lexicon, utilizing a lexical decision task format that can be delivered via online test administration. Together, this encompasses the major stimuli types that focus on different psycholinguistic grain sizes (Ziegler & Goswami, 2005) and cognitive processes used when reading texts for understanding in a task format that has demonstrated validity historically as an indicator of decoding (García & Cain, 2014). The comprehensiveness of the task is especially important when attempting to measure decoding in older students, where computationally complex, automated lexical processes may be called upon during the reading of a text: to recognize or access known words (in the mental lexicon of one's reading vocabulary), to decode novel or unknown words (creating a phonological representation to self-teach new words [Share, 1995]), or to associate spelling to words only previously heard before (i.e., match spelling to sound of known words).

**The current studies**

As reviewed above, reading theories such as the SVR, the LHQ and the STH provide different perspectives regarding the role decoding plays in reading comprehension. While the SVR implies linear relation between decoding and reading comprehension, the LHQ suggests the possibility of a lower bound decoding threshold below which the relation between decoding and reading comprehension becomes unpredictable. In the light of the STH, the existence of a lower bound decoding threshold could help explain why the self-teaching mechanism fails in certain groups of developing readers. Although all these theories have received support from empirical studies, these studies were conducted 1) with assumptions derived from different theoretical frameworks that 2) resulted in different conceptualizations and operationalizations of decoding

3) with a variety of samples that differed in age/grade and ability. The differences in the existing literature have created great complexities in understanding research findings on this topic. The goal of the current studies is to clarify theories and findings related to the role of decoding in reading comprehension across development by an examination of the Decoding Threshold Hypothesis, which posits that the relation between decoding and reading comprehension can only be reliably observed above a certain decoding threshold. We believe that the Decoding Threshold Hypothesis can integrate the different perspectives raised in the SVR, the LQH and the STH, with respect to low ability, older students. While the role of decoding in reading comprehension follows the SVR when students' decoding is above the decoding threshold, the identification of the decoding threshold itself is consistent with predictions of the LQH. Furthermore, students' being above and below the decoding threshold should have consequences for their reading development in the following years, a prediction consistent with the STH.

Specifically, this paper reports our exploration of two questions to test the Decoding Threshold Hypothesis. First, we explore whether the relation between decoding and reading comprehension is uniform across the continuum of reading ability. We hypothesize a lower bound decoding threshold, below which the relation between decoding and reading comprehension is weak and unpredictable, and above which a positive relation becomes apparent. This hypothesis is tested in Study 1. Second, we further explore whether being above or below the decoding threshold has consequences for students' future development in reading comprehension. We hypothesize that students who are below the decoding threshold should show less reading comprehension growth than students above the decoding threshold.

For the data collection, we collaborated with a school district in a mid-sized city on the US East coast and collected data from the public schools in the city. The school district has a large number of schools in an urban setting, with the great majority of students being African American and suffering from poverty. For Study 1, we were able to recruit more than 11,000 students who were across Grade 5-10, a development period for which some teachers may assume that decoding skills are mastered. Data provided by the school board show that the city had about 83,000 students from K-12 in the 2016-2017 academic year. Assuming there are roughly equal number of students at each grade level, Study 1 included about 30% of the local student population who were in Grade 5-10. Study 2 was a longitudinal study, and over 33,000 students who were from Grade 5-9 during their first participation participated in the study over the course of 3 years. This included the majority of the student population in Grade 5-9. It is worthwhile noted that the prevalence of non-proficient readers (based, for example, on state level tests) is higher than one might find in a nationally representative study. Thus, there is over-representation of students who may still struggle with decoding and comprehension than nationally representative studies of 5$^{th}$ to 9$^{th}$ grade students.

## Study 1: identifying decoding threshold with cross-sectional data

The goal of Study 1 was to examine the relation between decoding and reading comprehension in a sample of over 11,000 students from Grade 5 to 10. We examined whether the relation was indeed uniform across the ability distribution and, if not, whether there was a lower bound threshold of decoding ability, before it could predict reading comprehension. For this purpose, we used non-linear modeling, including quantile regression, classification performance analysis using the receiver operating characteristic (ROC) framework, as well as broken-line regression.

### Methods

**Participants.** Participants came from public schools in a mid-sized city located on the east coast of U.S.A. The data collection was in collaboration with the school board of the public schools in the city. The participating schools used the assessments as part of their curriculum, so the studies were exempt from Institute Review Board regulation.

A total of 11,765 students ranging from Grade 5-12 participated in the study and 11,000 of them completed a computerized reading battery of six subtests. Multivariate outliers were identified using the Mahalanobis distance (Stevens, 1984) and the top 5% cases with the highest Mahalanobis distance were excluded, thus resulting in a total of 10,450 valid cases. We replicated all analyses without removing the outliers, and the results did not alter any conclusions reported here other than slightly increasing standard errors.

After removal of outliers, the number of students in Grade 5-10 is provided in Table 1. Grade 11 and 12 had relatively small samples with 10 and 39 students respectively. Gender information for 3,647 students among all the students who had valid data (about 35%) was obtained after the data collection, with 1,879 female (52%) and 1,768 male (48%). Additionally, 9,266 of all the students (about 89%) had their race information available to the researchers, with 86.0% identifying themselves as African American, 12.3% White, 1% Asian, 0.2% American Indian/Alaskan Native, 0.2% Native Hawaiian/Other Pacific Islander, and 0.1% more than one race. The fact that the majority of our sample was African American students was consistent from the local student population: based on the school district website at the time of writing this paper, 80% of the students in the school district were African American.

**Materials.** Students took the Reading Inventory and Scholastic Evaluation (RISE) battery of reading tests (Sabatini, Bruce, Steinberg, & Weeks, 2015). RISE contains six subtests, each addressing a separate component of reading skills, including word recognition and decoding

(WRD); vocabulary; morphology; sentence processing; efficiency of basic reading comprehension; and reading comprehension (RC). The current paper focuses on two of these six subtests: WRD and RC. In the WRD subtest, students were presented with real words or non-words and they needed to decide whether each item is: 1) a real word, 2) not a real word, or 3) sounds exactly like a real word. In the RC subtest, students read passages and answer multiple-choice questions that have three options. The questions tapped students' ability to locate information in the text as well as their ability to draw inferences. Students were given sufficient time to complete the WRD and RC subtests, and their scale score on the two subtests were calculated from accuracy on each item without considering their speed.

The RISE has multiple forms that have been equated and linked, utilizing item response theory models (see Sabatini et al, 2015). It has been used with students ranging from Grade 4-12 with online test administration, in which students completed the six subtests in the above-mentioned order. The whole procedure took about 45 to 60 minutes. To compare performance across different forms/grades, there are unidimensional scaled scores of each subtest, which has a mean score of 250 and standard deviation of 15. Across these grade levels, reliability as reflected by Cronbach's alpha of the WRD subtest ranged from .90 to .92; for the RC subtest it ranged .72-.83 (Sabatini et al., 2015).

For criterion validity, in an unpublished study performed by the authors that included 542 students from Grade 4 to Grade 8, the correlations between RISE WRD to Gates-MacGinitie Reading Tests (GMRT; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000) Vocabulary and Comprehension subtests were $r = .726$ and $r = .636$, respectively; the correlation between RISE RC to GMRT Vocabulary and Comprehension subtests were $r = .669$ and $r = .746$ respectively. In another unpublished study performed by other researchers that includes 327 students from Grade 5, 7, 9 and 11, the correlation between RISE WRD raw score and TOWRE (Test of Word Reading Efficiency; Torgesen, Wagner, & Rashotte, 2012) was $r = .58$. In the same study, the correlation between RISE WRD raw score and TOSREC (Test of Silent Reading Efficiency and Comprehension; Wagner, Torgesen, Rashotte, & Pearson, 2010) for Grade 5 was $r = .44$, for Grade 7 was $r = .57$.

**Data Analysis.** To study the nonlinear changes in the relation between decoding and reading comprehension, we employed a series of analytical methods to answer this question, starting from testing nonlinearity as a weaker form of the hypothesis, then proceeding to explore the exact shape of the nonlinear relationship, the results of which led to a test of the stronger form of the hypothesis, which was the identification of a decoding threshold.

We first used quantile regression (Koenker, 2005) to examine whether the relation between decoding and reading comprehension was conditional on reading comprehension ability. Quantile regression is different from linear least squares regression in that instead of assuming a constant regression slope, it calculates several regression slopes at pre-defined quantiles of the dependent variable. Quantile regression has been widely used in economics and recently several educational researchers have also applied this method (e.g. Petscher & Kim, 2011). In our study, we performed quantile regression using WRD to predict RC scores at different RC quantiles to evaluate how the relation between WRD and RC changed across RC performance levels. We estimated the regression slopes of quantile regression at five RC quantiles (the $10^{th}$, $25^{th}$, $50^{th}$, $75^{th}$ and $90^{th}$ percentiles of RC) and compared them to the constant regression slope of linear regression. If some of these regression slopes obtained from quantile regression were significantly different from linear regression, it would suggest that the relation between decoding and reading comprehension is not uniform across the ability distribution.

Although quantile regression can reveal nonlinearity, it does not provide information about the exact shape of the nonlinear relationship. To address this problem, we proceeded with the ROC framework, using WRD score to predict low performance in RC, while varying the criterion of low RC performance. If the relation between WRD and RC is constant, then the prediction performance as revealed by ROC should remain constant as we vary the cutoff point for low RC performance. Conversely, if the prediction performance using WRD score to predict low RC changes as we vary the criterion for low RC, then it would suggest that the relation between WRD and RC changes. Furthermore, by looking at how the prediction performance changes as a function of RC cutoff point, we can make inferences about the shape of the nonlinear relation. We used the Area under the Curve (AUC) of ROC as the indicator of prediction performance. AUC is a measure that combines sensitivity and specificity and it has been shown to be a good indicator of the discrimination of prediction (Robin et al., 2017).

Finally, based on results obtained from quantile regression and ROC analysis which suggested there was one significant slope change in the relation between WRD and RC, we implemented broken-line regression. Broken-line regression is an extension to linear regression. Instead of estimating a single slope as in linear regression, broken-line regression estimates two different slopes divided by a certain point that is often referred to as the “broken” point (Muggeo, 2008). In a graphic representation, broken-line regression is represented by a broken line with two slopes, or two lines connected at the broken point. The broken line indicates that the relation between two variables changes at this broken point, which is often referred to as the threshold. We used a recently developed R package *lm.br* to explore whether WRD and RC has a broken line relationship (Adams, 2014). That is, if there exists a threshold below which the regression slope of RC on WRD is zero and above which the slope is significantly different from zero. The lm.br package can test this relationship, and if the relationship holds it also calculates a confidence interval of the position of this threshold for error estimates.

## Results

Table 1 shows the mean performance on WRD and RC as well as the correlation of the two measures separately for each grade as well as the whole sample. Students in higher grade levels generally showed higher scores on the two measures, except for students in grades 11 and 12 where the sample size was much smaller than other grade levels.

**Quantile Regression.** Table 2 shows how the regression slope of RC on WRD changes as a function of RC scores. In this table, we selected five RC percentiles: $10^{th}$, $25^{th}$, $50^{th}$, $75^{th}$ and $90^{th}$ and calculated the regression slopes at each of these five points. Compared to the constant regression slope estimated by linear regression, quantile regression shows that the regression coefficient is smaller at lower RC scores and it increases with RC scores. In other words, the regression slope becomes steeper at higher RC percentiles.

**ROC Analysis.** Quantile regression showed a trend indicating the relation between WRD and RC is weaker at lower RC scores. In an effort to have a closer look at how the strength of association between WRD and RC changed as a function of RC levels, we used ROC analysis. Specifically, we looked at how well WRD can predict high vs. low performance on RC. In other words, if we draw a cutoff point in RC and divide students into high and low performing RC groups, how well can their WRD score predict their group membership? Better categorization in this case suggests a stronger relation.

To measure classification performance, we followed the ROC framework and calculated the Area under ROC Curve (AUC). Figure 1 demonstrated how the AUC changes as a function of RC cutoff point with the grey band showing 95% confidence interval of AUC. Classification performance remained relatively low and had a large error when the RC cutoff point was set below 230, and it jumps when the RC cutoff point was between 230 and 240, and it reaches a plateau after we set the RC cutoff point at 250. In Figure 1, the jump in AUC between 230 and 240 indicates a qualitative change in the relation between WRD and RC at this area. Additionally, the plateau when RC cutoff point was set above 250 shows additional qualitative changes in this relationship are unlikely at this level.

**Threshold Analysis.** To confirm the "qualitative" change in the relation between WRD and RC as revealed by ROC analysis, we explored whether we can identify a WRD cutoff point such that below this point WRD does not predict RC but above this point WRD significantly predicts RC. Specifically, we varied the cutoff point in WRD score and repeatedly ran linear regression analysis between RC and WRD for groups below and above the WRD cutoff score, separately for each grade level (except Grade 11 and 12 due to the small sample size). We found that the hypothesized cutoff point in WRD indeed existed for each grade. Thus, the threshold is not merely a function of maturation as defined by grade level or extreme differences in students' age. These results are provided in Table 3.

Because the lower WRD group identified by linear regression has a smaller sample size compared to the higher WRD group, questions remain about whether the non-significant relation in the lower WRD group was due to insufficient statistical power. To deal with this problem, we used the broken-line regression described by Adams (2014) to identify the cutoff point and calculate the confidence intervals of this cutoff point. Because in broken-line regression, statistical significance testing is performed to examine whether there is a slope difference along the distribution of the independent variable, and the identification of the cutoff point is based on an examination of whether the 95% confidence interval of the identified cutoff point includes the lower or upper limit of the sample distribution, this method is not affected by unequal sample sizes of the low and high groups. Results of broken-line regression are reported in Table 3 in the first column on the right. The threshold WRD score was replicated by broken-line regression.

**Discussion**

Results of Study 1 replicates the magnitude of correlation between decoding and reading comprehension as reported by previous studies of older students. In our sample of Grade 5 to 12 students, the correlation between decoding and comprehension was .55 (Table 1). In comparison, according to the meta-analysis provided by García and Cain (2014), the correlation of the two measures for the same age group (older than 10 years old) was .47. Interestingly, if we only focus on the students who were above the decoding threshold, the correlation was .48 (Table 1), almost identical to García and Cain (2014).

We note that some readers may find it counter-intuitive that the strength of the correlation between decoding and comprehension is larger in older students (Table 1), when it is widely reported in the developmental reading literature that the relation of decoding to comprehension decreases across age/grades (e.g., Foorman, Francis, Shaywitz, Shywitz, & Fletcher, 1997; García & Cain, 2014; Vellutino, Tunmer, Jaccard, & Chen, 2003). With respect to this result, we note that our sample of students starts at grade 5 (about 10 years old) and above. Thus, we are already focusing on an “older” group of students, where the correlation between decoding and reading comprehension has already dropped in magnitude relative to elementary grade levels. In fact, the magnitude of correlations between decoding and comprehension we found are similar to those reported in García and Cain (2014). While some have reported a continued decline the strength of correlation after 5th grade (age 10), the decline is typically not steep. For example, Foorman et al., (1997) reported a decrease in correlation from grade 5 to grade 9 is no larger than .07: students from grades 5-7 had stable correlations of .70, .69 and .69 respectively; and students from grades 8 and 9 had correlations of .63 and .63. Thus, many studies report that the strength of correlation between decoding and reading comprehension only show significant decreases in earlier grade levels (e.g. grade 1 through grade 4) and that in older grade levels the relation stabilizes (also see García & Cain, 2014). This is consistent with our findings for older students.

In addition, the sample of students in this study is probably different from the samples in prior studies. The district sampled in this study has a high proportion of struggling readers attending low income schools. Thus, a large proportion of students are still struggling with mastery of grade level decoding skill. Consequently, the variability in this skill is still strongly related to their comprehension.

The key finding of Study 1 was that the relation between decoding and reading comprehension was non-linear. This non-linear relationship was replicated with various statistical methods, including quantile regression, classification analysis using WRD to predict RC ability group, and broken-line regression. These analyses revealed a threshold for students' decoding skills, below which decoding and reading comprehension was essentially unrelated. These findings extend our understanding of the role that decoding plays in reading comprehension. For students who had higher decoding and reading comprehension, it is true that their decoding is positively related to their reading comprehension, which supported LQH and SVR (Gough & Tunmer, 1986; Perfetti, 1988; Perfetti & Hart, 2002); however, for students who had lower decoding ability, comprehension is only weakly related to decoding. According to Table 3, as many as 38% students in Grade 5 and 19% students in Grade 10 belonged to this low decoding group, whose decoding performance did not show valid prediction to their reading comprehension. Importantly, the threshold was evident within each grade level. Thus, grade level cannot fully explain the incidence of a decoding threshold.

The threshold of WRD score suggests that decoding interventions aimed at improving students' reading comprehension might not lead to immediate progress for students below the WRD threshold, if the amount of intervention is not sufficient to push these students above the threshold. Since WRD is not correlated with RC below this WRD threshold, improvements in WRD might not lead to improvements in RC until the WRD threshold has been reached (but still not guaranteed). This is not to say that decoding interventions should not be implemented for students below the threshold, but rather that improvements in comprehension may not be immediately evident until the threshold is crossed (and even then one might predict that the relationship would establish itself only over time with practice reading). It remains to be investigated by future studies whether and what types of decoding intervention are effective, but our finding of the decoding threshold provides a necessary condition for the effectiveness of such intervention. To help inform future intervention studies, in Study 2 we look at a longitudinal dataset to see how students who are above and below the WRD threshold grow in RC in the following years.

## Study 2: Examining the decoding threshold with longitudinal data

In the light of the self-teaching mechanism (Share, 1995), the decoding threshold identified in Study 1 suggests that among students who are below the decoding threshold, an advantage in decoding does not translate into better reading comprehension. This implies that the self-teaching mechanism (Share, 1995) may be dependent on the decoding threshold, below which students' limited decoding ability negates the self-teaching mechanism that is necessary for word learning, which in turn is essential to growth in text comprehension (Quinn, Wagner, Petscher, & Lopez, 2015). The goal of Study 2 is to test this hypothesis, that is, whether students who are below the decoding threshold show no progress in their reading comprehension in the following years. Examining this question not only provides theoretical support for the self-teaching mechanism, but also has important implications for reading instruction. If the hypothesis is supported, students who continue to have severe decoding difficulties in grade five and beyond should be more likely to have stagnant growth in their reading comprehension later in development.

**Method**

**Participants.** The sample in this study was 34,016 students who came from the same school district as Study 1. These students participated in at least one of the four waves of data collection that happened in the fall semesters between the years 2011 through 2014; contributing to 55,863 complete test administrations between 2011 and 2014, with 17,721 students taking one wave of the test, 11,337 students taking two waves, 4,364 students three waves, and 594 four waves. The grade level distribution when students first participated in the study was: 8,143 commencing in Grade 5; 10,122 in Grade 6; 5,850 in Grade 7; 4,863 in Grade 8; and 5,038 in Grade 9. Note that some students who participated in Study 1 also participated in Study 2, but the data for Study 1 was collected on a different occasion and the data from Study 1 and Study 2 had no overlap.

A total of 18,666 students from the whole sample (55%) had their gender information available to our research team. This subsample consisted of 49% female and 51% male. In addition, 27,787 students from the whole sample (82%) had their race information available: 86% reporting as African American, 12% White, 1.1% Asian, 0.3% American Indian/Alaskan Native, 0.2% Native Hawaiian/Other Pacific Island, and 0.2% more than one race. Similar to

Study 1, the racial composition of our sample was comparable to the district level student demographics.

**Materials.** Students were administered the same RISE battery of tests as described in Study 1. Because RISE has multiple forms, efforts were taken to avoid students taking the same test form in different waves of data collection. Each subtest is linked across forms through a vertical IRT scale, hence, students' scores can be compared across different waves regardless of the form administered (Sabatini et al., 2015). Similar to Study 1, the current study focused on students' decoding (WRD) and reading comprehension (RC) performance.

**Analysis.** Growth curve modeling was applied to the longitudinal dataset (Rogosa, Brandt, & Zimowski, 1982). Level 1 models students' reading comprehension scores measured at different time points. Level 2 models characteristics of the student, including their decoding ability and grade level. Students' reading comprehension (RC) score was the dependent variable. The timing variable (*Time*) was the number of years since students first took the test, which varies from 0 (students who participated in the data collection in 2011) to 3 (e.g. those 2011 students who also participated in 2014). The second independent variable *Grade* was students' grade level during their first participation (time-invariant variable), and this variable is centered such that at fifth grade the variable *Grade* is set to 0. The third variable, *Decoding*, is students' decoding level. Because Study 1 revealed, with the cross-sectional data, a threshold score of 235 that remained relatively stable across different grade levels, in the current study we divided students into two decoding groups based on their decoding score: students in the below-threshold decoding group had WRD scores below 235, and the above-threshold decoding group had WRD scores above 234 (i.e. $\geq$235). The low decoding group had 11,403 students, and the high decoding group had 22,773 students.

Following conventions of longitudinal data analysis (Singer & Willett, 2003), we ran a number of models progressively, adding a new variable to the previous model at each step. By comparing the new model with the previous model, we evaluated how the addition of the new variable improved the model fit. Because not all students participated in all waves, we followed Singer and Willett (2003) and estimated the cohort effect using all students, and longitudinal effect using students who participated multiple waves of data collection. In doing so, the effect of time estimated by treating the data collections as variably spaced measurement occasions.

We first ran the unconditional mean model, Model A, where the intercept (initial status in RC score) was allowed to randomly vary. Model B is the unconditional growth model (Singer & Willett, 2003). In Model C, we added the *Grade* variable to students' initial status in RC. In Model D, the effect of grade level on the growth rate of RC was added and estimated. In Model E, the effect of decoding level on the initial RC status was added and estimated, and finally in Model F the effect of decoding level on the growth rate of RC, and the interaction between decoding group and grade level on both the initial status and growth rate of RC were added and estimated. For more details, we invite readers to refer to the equations of Models A-F under Table 4. The analysis was performed with R using the package *nlme* (Pinheiro, Bates, DebRoy, & Sarkar, 2017).

## Results

Results of the longitudinal analysis are summarized in Table 4. Model A revealed that the grand mean of students' RC performance across all individuals and all measurement occasions was 246. Model B, the unconditional growth model, showed that on average, students had an initial RC score of 246 at their first participation and averaged 1.86 points of improvement in RC score every year. By considering the linear effect of *Time*, 11% of the within-person variance was explained.

In Model C, we added the effect of *Grade* on the initial RC score. The model showed that an average Grade 5 student is expected to score 240 on the student's first RC test, and this value is expected to increase by 3.33 points in students who are one grade higher. For example, an average Grade 6 student is expected to score 240+3.33=243.33 in their first RC test. By adding the *Grade* variable, 9% of unexplained variation in students' initial RC score in the unconditional growth model was explained.

Model D further included the effect of *Grade* on the growth rate of RC score. Again, an average student from Grade 5 was estimated to have an initial RC score of 241, and each grade level increase added another 3.29 points to this initial RC score. For growth rate, an average student from Grade 5 was estimated to show 1.61 points of annual increase in RC. However, the variance components showed that Model D did not account for any unexplained variance in Model C. Additionally, the effect of grade level (3.29 points), which can serve as a cross-sectional estimate of students' RC annual growth, showed a large discrepancy compared to the longitudinal estimated effect of annual growth (1.61 points).

In Model E, we added the effect of decoding ability on students' initial RC scores. Grade 5 students who had normal decoding (e.g. WRD not lower than 235) had an average initial RC score of 249, with each grade level increase leading to 1.99 points higher in this initial RC score. In contrast, students whose decoding was below threshold were estimated to have much lower initial RC scores, about 16 points lower (i.e. more than 1 SD on the RISE norm) than their peers from the same grade level. Compared to Model D when decoding ability was not considered, 22% of unexplained variation in students' initial RC score was explained by Model E.

Finally, the effect of decoding group on the growth rate of RC was estimated in Model F. According to results of Model F, Grade 5 students who had normal decoding ability averaged 247 on their initial RC test. Each grade level was related to 2.78 points higher in initial RC. Interestingly, if a Grade 5 student was below the decoding threshold, each grade level up only corresponded to 0.36 (calculated as 2.78 – 2.42 based on Table 4) point increase in RC. This is the cross-sectional estimate of the effect of decoding threshold on the development of RC.

For the growth rate, Grade 5 students with normal decoding were estimated to have annual increase of 2.91 points in their RC performance, and each grade level increase was associated with .38 point more growth. In other words, the growth rate of RC was accelerating in these students. In contrast, Grade 5 students whose decoding ability was below the threshold had significantly lower initial RC scores and almost no annual growth in RC. Their initial RC score was estimated to be 235 (i.e. =247-11.84), and their annual RC growth rate was only .57 point (=2.91-2.34). Additionally, the acceleration in reading growth rate we observed in the normal decoding group was nullified in the low decoding group by the negative interaction between starting grade and decoding on the growth rate (.38-.40=-.02). Compared to Model E, adding the effect of decoding ability to growth rate explained about 15% of unexplained variation in growth rate.

**Discussion**

The main goal of Study 2 was to examine whether students' decoding status had consequences on their reading comprehension development. We hypothesized that students whose decoding was below the decoding threshold would not make any progress in their reading comprehension in the following years. To test this hypothesis, we divided students into below vs. above decoding threshold groups and tracked their reading comprehension performance on RISE over a period of three years. Longitudinal analysis indicated that RISE successfully captured students' RC development. In general, Grade 5 students above the decoding threshold had an average 2.91 points increase after a year. This estimate matched well with cross-sectional data showing the effect of students' initial grade level on RC score- each grade level increase was found to be associated with 2.78 points higher in RC score. The convergence between longitudinal estimates and cross-sectional estimates provides strong support to the validity of the longitudinal modeling, as well as the RISE assessment to track changes in reading development.

For students who were above the decoding threshold, their RC development showed acceleration across the grades. Grade level positively predicted their growth rate in RC, with each grade level corresponding to .38 point more annual increase in RC. To put this in context, Grade 5 students (above decoding threshold) were predicted to gain 2.91 points in RC one year later; for Grade 9 students, their RC improvement during the following year was predicted to be 2.91+.38*4=4.4 points.

Compared to the normal decoding group, the below threshold decoding group showed very different patterns in their RC scores and development. First, very poor decoding was associated with a low ceiling in RC performance. This is consistent with the widely accepted notion that decoding ability plays an important role in reading comprehension. Second, below threshold decoding was related to minimal growth in RC in the subsequent years. This provides empirical support to the STH (Share, 1995), which argues that decoding drives a self-teaching mechanism for word learning (vocabulary growth), which subsequently supports reading comprehension growth (Quinn et al., 2015). More importantly, the current study has successfully identified a lower bound decoding threshold below which almost no RC improvement was observed in later years. Students whose initial decoding score was below 235 on the RISE WRD test only had .6 point annual growth, compared to 2.9 of normal decoding students at Grade 5. Thus, this study contributes to the STH by providing a measurable decoding threshold below which the self-teaching mechanism is essentially inoperable. Finally, students with below-threshold decoding levels did not show any developmental acceleration in their RC which occurred in students with normal decoding.

These clear differences between below threshold decoding level students and normal decoding students provide further validation of the Decoding Threshold Hypothesis we evaluated in Study 1. The fact that the relation between decoding and RC changes as a function of decoding ability has not been considered in prior studies, with a few exceptions (e.g. O'Reilly, Sabatini, Bruce, Pillarisetti, & McCormick, 2012). Failure to consider the decoding threshold not only results in an inaccurate understanding of the relation between decoding and RC, but also leads to inconsistent estimation of students' RC development. In the Model D of our longitudinal analysis, in which decoding group was not considered, a mismatch was found between the cross-sectional estimate and the longitudinal estimate of students' annual RC growth. It is reasonable to believe that such inconsistencies will lead to significant confusions when researchers try to understand students' RC development with different developmental approaches (i.e. cross-sectional or longitudinal), had we not been aware that the development trajectory is dependent on students' decoding ability.

## General Discussion

While the authors of prominent reading theories agree on the important role of decoding in reading comprehension, they differ in predicting how exactly the two constructs are related. For example, the SVR, which viewed reading as a product of decoding and linguistic comprehension, assumes that decoding is linearly related to reading comprehension when linguistic comprehension is treated as a separate construct (Gough & Tunmer, 1986). In contrast, the LQH proposed a more dynamic relation between decoding and reading (Perfetti & Hart, 2002). According to the LQH, difficulties with decoding also interferes with higher level cognitive processes, both of which collectively affect reading comprehension. In other words, higher processes cannot operate when decoding is low, and as a result reading comprehension remains limited and cannot be predicted by decoding at this level. While the SVR and the LQH seem to be in conflict on the relation between decoding and reading comprehension, our hypothesis about the decoding threshold could be construed as a resolution to this conflict by defining a working range for the SVR and the LQH. Above the decoding threshold, decoding and reading comprehension are linearly related, following the SVR; below this threshold, the linear relation between decoding and reading comprehension disappeared, which is consistent with predictions from the LQH.

An interesting discovery of Study 1 was that the decoding threshold remained relatively stable across the grade levels. In other words, we found a constant decoding threshold for each grade level. This provided support for the robustness of the decoding threshold. Regardless of grade level, a decoding score below the decoding threshold almost always predicts low reading comprehension. Thus, the decoding threshold is not just a function of grade level, but rather a function of the level of students' decoding ability. When a student is above the fourth grade, it doesn't necessarily mean they have adequate decoding ability. In fact our data show that many students in the $10^{th}$ grade had inadequate decoding skills.

Since reading comprehension is no longer predicted by decoding below the decoding threshold, it appears that the whole burden of reading comprehension then rests solely on linguistic comprehension, according to the SVR ($R = D \times C$). In such a case, what drives reading comprehension? When examining students whose decoding ability is below the threshold, should we modify the SVR equation to include other compensatory factors such as prior knowledge, metacognition, inference making and contextual guessing to explain the reading performance? Although we do not have data on linguistic comprehension or other related measures, results of the longitudinal analysis show that when decoding is below the threshold, little if any development of reading comprehension is observed in the following years (i.e., students are probably not able to compensate with other linguistic or language abilities for severely weak decoding skills). In other words, nothing seems to be effectively driving reading comprehension when decoding is insufficient. Therefore, we do not propose a new equation to replace the SVR when decoding is low; instead our findings simply help define a working range for SVR. That is, if we treat the decoding threshold as a "zero" point for the decoding factor in SVR, then the SVR equation still works.

In Study 2, we followed students who had varying levels of decoding ability, some above the decoding threshold, and others below. Results showed that fifth grade students who were initially below the decoding threshold showed minimal improvement of reading comprehension performance in the following years. In comparison, fifth grade students who were above the decoding threshold had about 3 points (or 1/5 SD) of RC score increase each year. Additionally, the longitudinal modeling also revealed a mild acceleration in RC growth as students moved up the grade levels (about .38 points more annual growth with each grade level increase, see Table 4). However, this acceleration did not happen to students who were initially below the decoding threshold.

The longitudinal results related to decoding threshold provide support to the STH (Share, 1995). Decoding not only positively predicted reading comprehension cross-sectionally, but also predicted the growth rate of reading comprehension longitudinally (presumably allowing students to accumulate lexical knowledge necessary to support comprehension of increasingly diverse range of academic text content across school years). The fact that students who were below the decoding threshold showed little progress in reading comprehension in the following years reveals a necessary condition for self-teaching to occur. Self-teaching only happens when the developing reader has enough decoding ability to begin with. The decoding threshold we identified provides a quantitative description for the lower limit for the self-teaching mechanism.

This finding has implications for instructional practice. If a student who takes the RISE decoding test has a decoding score that is below the threshold value, and the student is older than Grade 4, when decoding instruction is no longer emphasized by U.S. teachers (Chall & Jacobs, 2003), then it is extremely unlikely that the student will make significant progress in reading comprehension in the following years. In such cases, additional decoding intervention might be needed before we bring the student back on track for reading comprehension growth. Previous studies have found that the word recognition interventions are not as effective as comprehension interventions (e.g. Wanzek, Wexler, Vaughn, & Ciullo, 2010). In the light of our decoding threshold findings, decoding intervention should probably be targeted to those who are below the threshold to be more effective, and the benefit of decoding intervention probably takes time to manifest in reading comprehension, as our longitudinal analysis reveals. Importantly, it should be noted that this seems to apply to students in any grade level ranging from Grade 5 through 9, since our results showed that decoding threshold remained constant across the grades (Study 1) and having a decoding score below the threshold predicted little growth in reading comprehension across all these grades (Study 2).

**Limitations and Future Directions**

In this paper we report our preliminary exploration of the Decoding Threshold Hypothesis. While the results are promising, they need to be interpreted with caution. The discovery of the decoding threshold raises as many questions as it answers. Below, we point out some of the limitations to be addressed in future studies.

First, we need to consider the representativeness of our sample and how it helped to reveal the lower bound threshold in decoding. Although we had a large number of students and the sample accounted for a large portion of the local student population (over 30% of the population in Study 1 and the majority of the student population in Study 2), all the students were from the same school district. According to information released by the school district, as of 2017 about 81% of the students in the school district are African American, 8% White and 9% Hispanic/Latino. Many schools have the majority of students living in poverty, as reflected by the free and reduced price lunch program provided on the district's website. This student population might have contributed to the unexpected finding that the relation between decoding and reading comprehension showed a slight increase (rather than decrease as shown in prior studies) in older students, in that even among the older students, many had not achieved decoding proficiency and therefore there was still a strong relationship to comprehension. One might expect an upper bound, where high proficiency in decoding is no longer able to predict high level, nuanced measures of comprehension as measured in more advanced comprehension tests than administered here (LaRusso et al., 2016). Thus, it is necessary to test the Decoding Threshold Hypothesis with different student populations and different comprehension measures.

Second, we had high attrition rate in Study 2, a challenge faced by many longitudinal studies. Luckily, in the longitudinal data analysis we showed that the effects of time (longitudinal) and grade level (cross sectional) were consistent with each other, this seems to implicate that the data were missing at random—if a large group of high/low ability students missed one wave of data collection, the effect of time and grade level would likely differ. Future studies should replicate the longitudinal model with a more complete data collection procedure, although the sample size can be smaller.

Third, it would be interesting and important to test the Decoding Threshold Hypothesis with other measures of decoding and word recognition. For future replication studies, it should be kept in mind that the current study benefited from using IRT vertically scaled subtests of the RISE WRD and RC, which were designed specifically for use with middle grades students and were group administered and automatically scored. This provided not only a sufficiently sensitive measure for estimating growth longitudinally across grades 5 to 10, but also the collection of a large number of cases across grades in the sample. Future studies evaluating the effect of decoding threshold status on the development of reading comprehension will benefit from using vertical scales that can be applied across a wide range of the developmental ability spectrum.

Finally and perhaps most importantly, the decoding threshold needs to be validated in an experimental intervention study. In the current studies, we validated the decoding threshold by examining how it mediated the relation between decoding and reading comprehension with correlational data. Although the results indicate that below the threshold there is no effect of decoding on reading comprehension, they do not confirm a causal effect above the threshold. Therefore, future studies should implement experimental studies to evaluate how intervention in decoding affects reading comprehension, and in particular, whether it is the case that only when the intensity of decoding intervention is sufficient to raise students' decoding above the threshold can we observe any long term, longitudinal impact on comprehension.

## Conclusion

In conclusion, the current research demonstrates that the relation between decoding and reading comprehension is not linear within and across Grade 5 to 10. The relation changed at a particular point on the decoding scale, which we termed as the “decoding threshold”. Below the threshold, reading comprehension scores had a low ceiling and differences in reading comprehension performance were not predicted by decoding score. Longitudinal analysis further showed that students below the decoding threshold did not have growth in reading comprehension scores in subsequent years. Decoding skill is not typically measured in students beyond Grade 4, but the results here suggest that it may an important construct to monitor in struggling readers beyond the elementary years.

**Reference**

Aarnoutse, C., Van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing, 14,* 61-89. https://doi.org/10.1023/A:1008128417862

Adams, M. (2014). lm. br: An R Package for Broken Line Regression. http://cran.pau.edu.tr/web/packages/lm.br/vignettes/lm.br.pdf

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2008). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, *42*, 163-176. https://doi.org/10.1177/0022219408326219

Catts, H. W., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278-293. doi:10.1044/1092-4388(2006/023)

Chall, J. S., & Jacobs, V. A. (2003). The classic study on poor children's fourth-grade slump. *American educator*, *27*(1), 14-15.

Chen, R. S., & Vellutino, F. R. (1997). Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research, 29*, 1-24. https://doi.org/10.1080/10862969709547947

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences 2nd edn*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology, 82*, 185-199. https://doi.org/10.1016/S0022-0965(02)00008-5

Elliott, C. D., Smith, P., & McCulloch, K. (1996). *British Ability Scales (BAS II): administration and scoring manual*. London: NFER-Nelson.

Foorman, B.R., Francis, D.J, Shaywitz, S.E., Shaywitz, B.A., & Fletcher, J.M. (1997). The case for early reading interventions. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 243-264). Mahwah, NJ: Erlbaum.

Foorman, B. R., Petscher, Y., & Herrera, S. (2018). Unique and common effects of decoding and language factors in predicting reading comprehension in grades 1–10. *Learning and Individual Differences, 63*, 12-23. doi:10.1016/j.lindif.2018.02.011

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading, 5*, 239-256. 10.1207/s1532799xssr0503_3

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

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2,* 127-160. https://doi.org/10.1007/BF00401799

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review*, *34*, 9-26.

Houck, B. D., & Ross, K. (2012). Dismantling the myth of learning to read and reading to learn. *ASCD Express*, 7(11). Retrieved from http://www.ascd.org/ascd-express/vol7/711-houck.aspx

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing, 19*, 339-361. https://doi.org/10.1007/s11145-005-4644-2

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447. http://dx.doi.org/10.1037/0022-0663.80.4.437

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*, 281-300. https://doi.org/10.1080/10888430802132279

Kim, J. S., Hemphill, L., Troyer, M., Thomson, J. M., Jones, S. M., LaRusso, M. D., & Donovan, S. (2017). Engaging struggling adolescent readers to improve reading skills. Reading research quarterly, 52(3), 357-382. https://doi.org/10.1002/rrq.171

Kirby, J. R., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy*, *42*, 75-82. 10.1111/j.1741-4369.2008.00487.x

Koenker, R. (2005). *Quantile regression*. New York, NY: Cambridge university press.

LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., ... & Snow, C. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness*, *9*(2), 201-222. https://doi.org/10.1080/19345747.2015.1116035

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2000). *Gates-MacGinitie reading tests- 4th Ed. (GMRT-4).* Rolling Meadows, IL: Riverside Publishing.

Muggeo, V. M. (2008). Segmented: an R package to fit regression models with broken-line relationships. *R news, 8*, 20-25.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & C. M. E (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). New York, NY: Psychology Press.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70*, B1-B13. https://doi.org/10.1016/S0010-0277(99)00004-9

O'Reilly, T., Sabatini, J., Bruce, K., Pillarisetti, S., & McCormick, C. (2012). Middle school reading assessment: Measuring what matters under a RTI framework. *Reading Psychology, 33*, 162-189. https://doi.org/10.1080/02702711.2012.631865

Perfetti, C. A. (1988). Verbal efficiency in reading ability. In M. Daneman, G. E. Mackinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice*. San Diego, CA: Academic Press.

Perfetti, C. A., & Adlof, S. M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J. Sabatini, E. Albro & T. O'Reilly (Eds.), *Measuring up: Advances in how to assess reading ability* (pp. 3-20). Lanham, MD: Rowman & Littlefield Education.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. T. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (Vol. 11, pp. 67-86). Amsterdam, Netherlands: John Benjamins Publishing.

Petscher, Y., & Kim, Y.-S. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology, 49*, 107-129. https://doi.org/10.1016/j.jsp.2010.09.004

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2017). nlme: linear and nonlinear mixed effects models. Retrieved October 1, 2017, from http://CRAN.R-project.org/package=nlme

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child development, 86*(1), 159-175. doi:10.1111/cdev.12292

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., . . . Robin, M. X. (2017). Package 'pROC'. Retrieved October 1, 2017, from http://cvsup3.pl.freebsd.org/pub/mirrors/CRAN/web/packages/pROC/pROC.pdf

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological bulletin, 92*, 726-748. http://dx.doi.org/10.1037/0033-2909.92.3.726

Sabatini, J., Bruce, K., Steinberg, J., & Weeks, J. (2015). SARA Reading Components Tests, RISE Forms: Technical Adequacy and Test Design (ETS RR-15-32). Princeton, NJ: Educational Testing Service.

Sabatini, J., Sawaki, Y., Shore, J. R., & Scarborough, H. S. (2010). Relationships among reading skills of adults with low literacy. *Journal of Learning Disabilities, 43*, 122-138. https://doi.org/10.1177/0022219409359343

Sabatini, J., Wang, Z., & O'Reilly, T. (in press). Relating reading comprehension to oral reading performance in the NAEP fourth-grade special study of oral reading. *Reading Research Quarterly*.

Savage, R. (2006). Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. *Scientific Studies of Reading, 10*, 143-164. https://doi.org/10.1207/s1532799xssr1002_2

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review, 96*, 523-568. http://dx.doi.org/10.1037/0033-295X.96.4.523

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151-218. https://doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology, 87*, 267-298. https://doi.org/10.1016/j.jecp.2004.01.001

Share, D. L., & Shalev, C. (2004). Self-teaching in normal and disabled readers. *Reading and Writing, 17*, 769-800. https://doi.org/10.1007/s11145-004-2658-9

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford university press.

Stevens, J. P. (1984). Outliers and influential data points in regression analysis. *Psychological bulletin, 95*, 334-344. http://dx.doi.org/10.1037/0033-2909.95.2.334

Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading, 32,* 383-401.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *TOWRE-2 Examiner's Manual*. Austin, TX: Pro-Ed.

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The quarterly journal of experimental psychology*, *67*, 703-727. https://doi.org/10.1080/17470218.2013.826258

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. Scientific Studies of Reading, 11(1), 3-32. doi:10.1080/10888430709336632

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *TOSREC: Test of silent reading efficiency and comprehension*. Austin, TX: Pro-Ed.

Walczyk, J. J., Marsiglia, C. S., Johns, A. K., & Bryan, K. S. (2004). Children's compensations for poorly automated reading skills. *Discourse Processes, 37*(1), 47-66. https://doi.org/10.1207/s15326950dp3701_3

Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing, 23*(8), 889-912. doi:10.1007/s11145-009-9179-5

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin, 131*(1), 3-29. doi: 10.1037/0033-2909.131.1.3

Table 1. Descriptive analysis results by grade and decoding group.

| Grade | Decoding Group (Below or Above Threshold) | Mean WRD (SD) | Mean RC (SD) | Correlation (WRD, RC) | N |
|---|---|---|---|---|---|
| 5 | Below | 228(4) | 234(13) | .04 | 722 |
| 5 | Above | 249(10) | 247(19) | .43* | 1316 |
| 5 | All | 241(13) | 242(18) | .49* | 2038 |
| 6 | Below | 229(4) | 234(10) | .06 | 793 |
| 6 | Above | 249(10) | 247(19) | .44* | 1685 |
| 6 | All | 243(13) | 242(18) | .51* | 2478 |
| 7 | Below | 228(4) | 233(11) | .07 | 759 |
| 7 | Above | 252(12) | 249(22) | .45* | 1609 |
| 7 | All | 244(15) | 244(21) | .53* | 2368 |
| 8 | Below | 228(4) | 234(12) | .12* | 505 |
| 8 | Above | 254(13) | 253(26) | .45* | 1581 |
| 8 | All | 248(16) | 248(25) | .52* | 2086 |
| 9 | Below | 229(5) | 236(13) | .04 | 151 |
| 9 | Above | 259(14) | 261(30) | .47* | 746 |
| 9 | All | 254(17) | 257(29) | .54* | 897 |
| 10 | Below | 230(3) | 234(11) | .04 | 69 |
| 10 | Above | 258(14) | 260(27) | .50* | 465 |
| 10 | All | 255(16) | 257(27) | .57* | 534 |
| 11 & 12 | Below | 230(3) | 235(11) | -.10 | 11 |
| 11 & 12 | Above | 246(8) | 246(12) | .14 | 38 |
| 11 & 12 | All | 242(10) | 244(13) | .33* | 49 |
| 5 to 12 | Below | 228(4) | 234(11) | .06* | 3010 |
| 5 to 12 | Above | 252(12) | 251(24) | .48* | 7440 |
| 5 to 12 | All | 245(15) | 246(22) | .55* | 10450 |

*$p<.05$; WRD=word recognition and decoding; RC=reading comprehension. Copyright by Educational Testing Service, 2018. All rights reserved. Reprinted with permission.

Table 2. Regression coefficients of decoding predicting reading comprehension in quantile regression and linear regression.

| | Linear Regression (Assuming Constant Slope) | Quantile Regression RC at 0.1 Quantile | Quantile Regression RC at 0.25 Quantile | Quantile Regression RC at 0.5 Quantile | Quantile Regression RC at 0.75 Quantile | Quantile Regression RC at 0.9 Quantile |
|---|---|---|---|---|---|---|
| *B* | .81 | 0.31* | 0.42* | 0.69* | .97* | 1.24* |
| *SE(B)* | 0.012 | 0.017 | 0.013 | 0.013 | 0.015 | 0.025 |
| *t* | 67.01 | 19.08 | 31.96 | 52.07 | 65.16 | 49.56 |
| *df* | 10448 | 10448 | 10448 | 10448 | 10448 | 10448 |
| *p* | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 |

***** Regression coefficient significantly differ from linear regression, $p<.05$; RC=reading comprehension. Copyright by Educational Testing Service, 2018. All rights reserved. Reprinted with permission.

Table 3. WRD cutoff point identified by linear regression and broken line regression: below the cutoff point there was no relation between WRD and RC.

| *N* | Grade | Cutoff WRD score (Linear Regression) | Cutoff WRD percentile (Linear Regression) | Cutoff WRD (Broken Line Regression) | Cutoff WRD 95% CI (Broken Line Regression) |
|---|---|---|---|---|---|
| 2038 | 5 | 235 | 38th | 232 | [229, 235] |
| 2478 | 6 | 235 | 35th | 233 | [230, 236] |
| 2368 | 7 | 234 | 32th | 232 | [228, 235] |
| 2086 | 8 | 231 | 18th | 234 | [230, 238] |
| 897 | 9 | 237 | 20th | 233 | [226, 241] |
| 534 | 10 | 238 | 19th | NA | [-∞, 239] |

Copyright by Educational Testing Service, 2018. All rights reserved. Reprinted with permission.

Table 4. Longitudinal analysis evaluating factors related to reading comprehension development.

| | | | Model A | Model B | Model C | Model D | Model E | Model F |
|---|---|---|---|---|---|---|---|---|
| Fixed Effects | | | | | | | | |
| Initial status $\pi_{0i}$ | Intercept | $\gamma_{00}$ | 246* (.11) | 246* (.11) | 240* (.16) | 241* (.16) | 249* (.18) | 247* (.20) |
| | Starting Grade | $\gamma_{01}$ | | | 3.33* (.08) | 3.29* (.08) | 1.99* (.07) | 2.78* (.09) |
| | Decoding<235 | $\gamma_{02}$ | | | | | -16.07* (.20) | -11.84* (.31) |
| | Starting Grade by Decoding<235 | $\gamma_{03}$ | | | | | | -2.42* (.16) |
| Change rate $\pi_{1i}$ | Intercept | $\gamma_{10}$ | | 1.86* (.09) | 2.05* (.09) | 1.61* (.15) | 1.74* (.15) | 2.91* (.20) |
| | Starting Grade | $\gamma_{11}$ | | | | .40* (.12) | .35* (.11) | .38* (.14) |
| | Decoding<235 | $\gamma_{12}$ | | | | | | -2.34* (.31) |
| | Starting Grade by Decoding <235 | $\gamma_{13}$ | | | | | | -.40* (.22) |
| Variance components | | | | | | | | |
| Level-1 | Within-person | $\sigma_\varepsilon^2$ | 181 | 163 | 163 | 163 | 164 | 164 |
| Level-2 | In initial status | $\sigma_0^2$ | 264 | 236 | 215 | 216 | 163 | 160 |
| | In rate of change | $\sigma_1^2$ | | 13.6 | 13.4 | 13.4 | 13.6 | 11.5 |
| Goodness-of-fit | | | | | | | | |
| | Deviance | | 488769 | 487157 | 485280 | 485267 | 479412 | 478939 |
| | AIC | | 488775 | 487169 | 485294 | 485283 | 479430 | 478963 |
| | BIC | | 488802 | 487222 | 485356 | 485354 | 479511 | 479071 |
| | Improvement compared to previous model ($\chi^2$) | | Baseline | 1612* | 1877* | 13* | 5854* | 473* |

Note: Standard Error in parentheses; * *p*<.01; Model A: $RC_{ij} = \gamma_{00} + \zeta_{0i} + \varepsilon_{ij}$

Model B: $RC_{ij} = \gamma_{00} + \zeta_{0i} + (\gamma_{10} + \zeta_{1i}) \times Time_{ij} + \varepsilon_{ij}$

Model C: $RC_{ij} = \gamma_{00} + \gamma_{01} \times Grade_i + \zeta_{0i} + (\gamma_{10} + \zeta_{1i}) \times Time_{ij} + \varepsilon_{ij}$

Model D: $RC_{ij} = \gamma_{00} + \gamma_{01} \times Grade_i + \zeta_{0i} + (\gamma_{10} + \gamma_{11} \times Grade_i + \zeta_{1i}) \times Time_{ij} + \varepsilon_{ij}$

Model E: $RC_{ij} = \gamma_{00} + \gamma_{01} \times Grade_i + \gamma_{02} \times Decoding_i + \zeta_{0i} + (\gamma_{10} + \gamma_{11} \times Grade_i + \zeta_{1i}) \times Time_{ij} + \varepsilon_{ij}$

Model F: $RC_{ij} = \gamma_{00} + \gamma_{01} \times Grade_i + \gamma_{02} \times Decoding_i + \gamma_{03} \times Grade \times Decoding + \zeta_{0i}$
$+(\gamma_{10} + \gamma_{11} \times Grade_i + \gamma_{12} \times Decoding_i + \gamma_{13} \times Grade_i \times Decoding_i + \zeta_{1i}) \times Time_{ij} + \varepsilon_{ij}$

Copyright by Educational Testing Service, 2018. All rights reserved. Reprinted with permission.

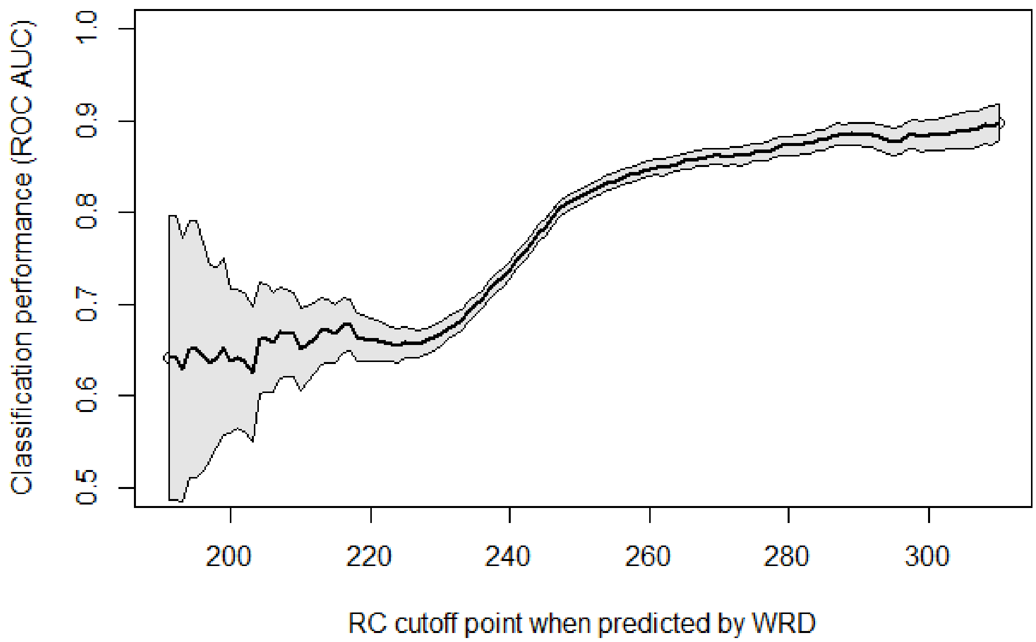


Figure 1. Performance of categorization using WRD to predict RC groups as a function of RC cutoff point; band shows 95% confidence interval. ROC=receiver operating characteristic; AUC=area under the curve; WRD=word recognition and decoding; RC=reading comprehension. Copyright by Educational Testing Service, 2018. All rights reserved. Reprinted with permission.