



# Measuring Elementary School Teachers' Knowledge of Teaching Vocabulary

Douglas M. Mosher  
Harvard University

James S. Kim  
Harvard University

Teacher Knowledge in the field of literacy has become a priority across the United States, with many states passing legislation requiring that all teachers receive adequate training on the Science of Reading. One essential component of literacy development and text comprehension is vocabulary acquisition. This study examines the psychometric properties of items from the Teacher Knowledge of Vocabulary Survey (Duguay et al., 2016) as well as 11 new items focusing on domain-specific vocabulary instruction. Using responses from 400 third-grade teachers, 2-parameter item response theory models revealed that on average, items were relatively easy for participating teachers. Multilevel models revealed that the measure did not significantly predict vocabulary or reading comprehension achievement for third-grade students ( $N = 6,977$ ). The discussion considers reasons why item-difficulty dropped from the initial administration as well as implications for future use and research.

VERSION: August 2025

Suggested citation: Mosher, Douglas M., and James S. Kim. (2025). Measuring Elementary School Teachers' Knowledge of Teaching Vocabulary. (EdWorkingPaper: 25-1264). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/syn8-f919>

Measuring Elementary School Teachers' Knowledge of Teaching Vocabulary

Douglas M. Mosher

James S. Kim

*Harvard University*

## Abstract

Teacher Knowledge in the field of literacy has become a priority across the United States, with many states passing legislation requiring that all teachers receive adequate training on the Science of Reading. One essential component of literacy development and text comprehension is vocabulary acquisition. This study examines the psychometric properties of items from the Teacher Knowledge of Vocabulary Survey (Duguay et al., 2016) as well as 11 new items focusing on domain-specific vocabulary instruction. Using responses from 400 third-grade teachers, 2-parameter item response theory models revealed that on average, items were relatively easy for participating teachers. Multilevel models revealed that the measure did not significantly predict vocabulary or reading comprehension achievement for third-grade students ( $N = 6,977$ ). The discussion considers reasons why item-difficulty dropped from the initial administration as well as implications for future use and research.

**Keywords:** vocabulary, science of reading, literacy, teacher knowledge, reading comprehension

## Introduction

Ample research over four decades has documented vocabulary knowledge as a significant predictor of reading comprehension (Elleman et al., 2009; McKeown et al., 1983; Stahl & Fairbanks, 1986; Wright & Cervetti, 2016). Yet while vocabulary knowledge among students is essential to developing proficient reading ability, far less is known about measuring teachers' knowledge of vocabulary and the association between that knowledge and student achievement. In the present study, we focus on a conceptual replication of an existing measure as well as an expansion of the measure by adding additional items. By conceptual replication, we mean that we designed this study to vary the features of the original study in order to extend findings and employ more comprehensive analyses (Bollen et al., 2015; Derksen & Morawski, 2022; Makel & Plucker, 2014; Nosek et al., 2015; Van Bavel et al., 2016). We examine the psychometric properties of items from three sections of the Teacher Knowledge of Vocabulary Survey (TKVS; Duguay et al., 2016) as well as 11 new items designed to measure teachers' content knowledge of teaching domain-specific vocabulary. We conclude our analyses by determining whether and to what extent teacher knowledge of vocabulary as a unitary construct is a significant predictor of students' reading comprehension ability.

### The Importance of Teacher Knowledge

Students learn by being in school, and consequently, rely on teachers to communicate and convey content knowledge in the domains of literacy, math, science, and social studies. In order to teach, however, educators must possess *content knowledge* and *pedagogical content knowledge* (Shulman, 1986). Content knowledge refers to possessing knowledge on a given topic or area of study (e.g., knowledge of phonemic awareness, vocabulary, orthography, etc.). Yet simply possessing content knowledge does not always translate to effective teaching. In that regard, *pedagogical content knowledge* pertains to effectively conveying specific knowledge to

learners. Thus, a teacher needs both content knowledge and pedagogical content knowledge to convey what they already know to their students who are in the process of learning. If the ultimate goal of teaching is to help students learn and acquire knowledge, then measuring teachers' content knowledge seems essential in understanding what teachers actually know and if teachers are being asked to teach that which they do not know (Peter Effect; Applegate & Applegate, 2004).

### **Recent Legislation Has Resulted in Professional Development on the Science of Reading for Teachers**

Within a decade, 32 states and the District of Columbia have passed laws requiring that reading instruction be tied to the Science of Reading (Schwartz, 2023). This legislation includes a number of different components: (a) requiring teachers participate in approved professional development (PD)/coursework, (b) using certain curricular materials, (c) types of instruction, and (d) assessments aligned with the science of reading. In some cases, legislation has required that teachers participate in relevant PD in order to receive their teaching certification or renewal of teaching licenses. Indeed, legislators are requiring teachers to acquire the necessary content knowledge about the science of reading with the hopes that this knowledge will improve literacy outcomes for students.

There is evidence that suggests there is a positive and significant association between increased content knowledge and student literacy outcomes. Mississippi became the first state to adopt science of reading legislation in 2013 (Schwartz, 2023), and in response to low National Assessment of Educational Progress (NAEP) scores, the Mississippi Department of Education instituted a professional development series – Language Essentials for Teachers of Reading and Spelling (LETRS) – targeting literacy practices such as “theoretical models from reading science,

phonology, basic and advanced phonics, screening and educational diagnostic assessment, as well as teaching vocabulary, language and reading comprehension, and writing” (Lexia, n.d., p. 7). Concurrent with LETRS training (and subsequently AIM Pathways), the Regional Educational Laboratory (REL) developed and launched the comprehensive Assessment of Teacher Knowledge (ATK) to measure early literacy skills. Results indicated significant positive associations between attending the professional development series and greater knowledge of literacy skills (Folsom et al., 2017). Interestingly, there was also a significant increase in NAEP scores for Mississippi students, suggesting that increased teacher knowledge might be a key factor in improving student outcomes (U.S. Department of Education, 2019), although there is no causal evidence to indicate that increasing teachers’ knowledge directly impacted student test scores, especially given the difficulty of isolating potentially confounding factors (i.e., child attending preschool).

Yet, one large-scale randomized controlled trial indicated that the LETRS training yielded mixed results on improving teachers’ knowledge about reading instruction. While there was a significant treatment effect on teachers’ total score of teacher knowledge, when Garet and colleagues (2008) decomposed the knowledge score into two domains – *word-level components* (i.e., phonemic awareness, phonics, and reading fluency) and *meaning-level components* (i.e., vocabulary and reading comprehension) – there was only a significant treatment effect on knowledge of word-level components. Findings suggest that participating in the LETRS training helped build teachers’ knowledge of foundational literacy skills but did very little to develop teachers’ knowledge of other essential aspects of literacy such as vocabulary and reading comprehension. Furthermore, there were no significant correlations between teacher knowledge and student outcomes (Garet et al., 2008).

Other studies, however, suggest that improved teacher knowledge is predictive of student outcomes. In one study, teachers who provided explicit instruction in reading and had substantial literacy content knowledge actually predicted increased word reading among students compared to students in classrooms taught by teachers with limited knowledge (Piasta et al., 2009). More broadly, teacher knowledge has also been found to alter classroom practice, which can then lead to improved student achievement (McCutchen et al., 2002). Another study found that students in classrooms of knowledgeable teachers had improved reading comprehension in first grade, but the significant association between knowledgeable teachers and students' reading comprehension achievement disappeared in second and third grade (Carlisle et al., 2011).

### **Measuring Teachers' Knowledge of General Vocabulary Instruction**

Broadly speaking, the term *teacher knowledge* can encompass many aspects of the science of reading. While some measures such as the Teacher Understanding of Literacy Constructs and Evidence-Based Instructional Practices (TULIP) survey cover five key domains of literacy (phonological awareness, phonics/decoding/encoding, reading fluency, oral language, and reading comprehension; Hall et al., 2024), other measures of teacher knowledge focus exclusively on single components of literacy. One such aspect, vocabulary knowledge, is essential for readers to comprehend texts successfully (Anderson & Freebody, 1981; Elleman et al., 2009; Stahl & Fairbanks, 1986), yet there have been only a handful of different measures examining teacher knowledge of vocabulary (Duguay et al., 2016; Phillips et al., 2019) and at the more granular level, morphology (Davis et al., 2021; McMahan et al., 2019). Duguay and colleagues (2016) developed the TKVS, the first measure solely dedicated to measuring teachers' knowledge of vocabulary and instruction, to include six sections: (1) development of vocabulary knowledge, (2) vocabulary and instruction: providing rich and varied language experiences, (3)

vocabulary instruction: teaching individual words, (4) vocabulary instruction: teaching word-learning strategies, (5) vocabulary instruction: fostering word consciousness, and (6) vocabulary instruction for English learners. The TKVS was validated using responses from 50 teachers, of which 35 taught in Northeastern middle schools and 15 taught second grade in a large metropolitan district in Texas. A Rasch model was used to evaluate each item's difficulty as well as correlating teachers' TKVS score with emergent, intermediate, and expert cut scores determined by three expert raters using the Angoff method. The authors indicated that the measure could be useful as a diagnostic tool in assessing potential topics to cover in PD sessions as TKVS items spanned the continuum from easy to difficult, and though the measure is promising given the wide-range of location parameters on the theta scale, there is no correlational evidence linking the instrument to student outcomes.

### **Measuring Teachers' Knowledge of Domain-Specific Vocabulary Instruction is a Relatively Unexplored Construct**

While the TKVS measures general knowledge of vocabulary instruction, an emerging area of focus around vocabulary instruction is instruction of domain-specific vocabulary (i.e., words found in the domains of science, social studies, etc.). To our knowledge, there are no measures that address teacher knowledge of domain-specific vocabulary instruction. Words that are tied to specific content areas represent kernels of knowledge that are essential to building domain and topic knowledge (Alexander, 2003), and domain-specific vocabulary words are an integral part of content literacy interventions (Kim et al., 2020; 2021; 2023; 2024a) and have been shown to mediate reading comprehension outcomes and writing outcomes (Mosher et al., 2024; Relyea et al., 2024). At the more granular level, examining vocabulary words, specifically domain-specific vocabulary, at the sublexical level has been a focus of recent content literacy

interventions (Kim et al., 2023). In his Lexical Quality Hypothesis, Perfetti (2007) argues that deep knowledge of a word requires learners to understand its form and meaning. Thus, focusing on morphology of vocabulary is essential to understanding the different components comprising words, which ultimately contributes to greater depth of understanding. Yet, teachers have been found to have less knowledge of morphology than of other domains such as decoding, encoding, phonological sensitivity, and phonemic awareness (McMahan et al., 2019). Recent measures of teacher knowledge have included items dedicated to assessing knowledge of morphology (Davis et al., 2021; McMahan et al., 2019), yet neither study examined correlations between the measures and student achievement. Indeed, in a review of 20 studies focusing on the science of reading in teacher preparation and training programs, the authors highlighted the need “to gain a more comprehensive picture of the influence of teacher knowledge on student literacy outcomes, especially reading comprehension” (Hudson et al., 2021, p. S311). While developing measures is an important marker of tracking teacher knowledge, the lack of evidence linking teacher-level measures to student outcomes leaves an incomplete picture in the research literature on the significance of teacher knowledge in promoting improved student outcomes and if it is a key mechanism worth investing resources.

### **Third-Grade Reading Achievement is a Key Indicator in Explaining Long-Term Outcomes**

The few studies that have correlated teacher knowledge with student outcomes have often focused on third-grade student outcomes (Carlisle et al., 2011; Garet et al., 2008). Third grade is the first year that students receive standardized end-of-grade assessments that measure reading comprehension ability using grade-level passages. Prior to third-grade, students are assessed with standardized measures (e.g., DIBELS), but those measures target key foundational skills like

phonemic awareness, fluency, and basic comprehension rather than deep comprehension. Research from the past two decades has documented that students who are considered at or above grade-level proficiency at the end of third grade are significantly more likely to attend college than students who are deemed below grade-level (Lesnick et al., 2010), and struggling readers are far less likely to graduate high school than their more proficient counterparts (Annie E. Casey Foundation, 2011).

In sum, there is compelling evidence to indicate that teacher knowledge is malleable, yet an inconsistent, albeit important predictor of student achievement. Development of measures focusing on specific elements tied to the science of reading – i.e., vocabulary, phonics, fluency – helps shed light on the importance of each component. While the TKVS has already been psychometrically evaluated (Duguay et al., 2016), it was done with mostly middle-school teachers and a small number of second-grade teachers. As third-grade is such a crucial year in measuring student achievement, for a grade ultimately evaluated by high-stakes assessments, measuring third-grade teachers' knowledge of teaching vocabulary – with vocabulary already established as a key predictor of reading comprehension – and correlating teacher knowledge with student achievement would seem to be important in assessing the construct validity of items from the TKVS and 11 additional items covering knowledge of domain-specific vocabulary and morphology, given that limited evidence exists to suggest teacher knowledge of literacy practices supports third-grade reading achievement (Carlisle et al., 2011). Furthermore, measuring teacher knowledge of teaching domain-specific vocabulary and the morphological components of domain-specific vocabulary words is an area that needs further attention given the emergence of numerous content literacy programs that privilege knowledge-building through the inclusion of

domain-specific vocabulary instruction (Connor et al., 2017; 2018; Kim et al., 2020; 2021; 2023; 2023; 2024; Guthrie et al., 2004; Guthrie & Klauda, 2014; Romance & Vitale, 2001).

### **The Present Study**

In the present study, we aimed to evaluate the psychometric properties of three sections of the TKVS as well as 11 additional items measuring teachers' knowledge of domain-specific vocabulary instruction. We conducted a conceptual replication in that we administered items from three out of the six sections from the TKVS measure, but we varied the features of the measure by forcing respondents to pick either *True* or *False* for each item, thereby eliminating the *I don't know* option after consulting with psychometricians. Furthermore, our analyses both directly replicate Duguay and colleagues' (2016) study and expand on it by employing a more comprehensive evaluation of TKVS items, given that we had a sample size eight times larger than the original study. We also append the TKVS by adding 11 multiple-choice items designed to measure teachers' knowledge of domain-specific vocabulary instruction and knowledge of teaching the morphological components of domain-specific vocabulary words. Using responses from 400 third-grade teachers and their respective students ( $N = 6,977$ ), we then correlated teacher performance on the revised and expanded measure with student performance on various vocabulary and reading comprehension assessments.

### **Research Questions**

1. What are the psychometric properties of items from three sections of the TKVS?
2. What are the psychometric properties of TKVS items and domain-specific vocabulary items?
3. To what extent do all items measuring teachers' knowledge of vocabulary instruction as a unitary construct predict students' vocabulary and reading comprehension achievement?

## Method

### Teacher Measures

We included selected items from the TKVS as well as 11 developed items focusing on domain-specific vocabulary instruction. We piloted the 11 items the previous year on 96 third-grade teachers and made substantial revisions based on our results.

#### *TKVS*

The TKVS includes six sections, and while all are important, we were particularly interested in three domains: (a) development of vocabulary knowledge, (b) teaching individual words, and (c) fostering word consciousness. Because the teacher knowledge section was part of a larger survey, we were mindful of survey length. Consequently, in addition to omitting three of the six domains entirely, we dropped three items from the development of vocabulary knowledge section. Omitted items were not relevant to the type of word-learning instruction covered in the Model of Reading Engagement (MORE) intervention, in which all teachers who completed the survey participated.

Original TKVS items had three answer choices: *True*, *False*, or *I don't know*. We deviated from the original measure in that we wanted to force teachers to pick an answer. As such, we removed the *I don't know* answer option after consulting with psychometricians. Items were scored dichotomously as either correct or incorrect.

#### *Applied Domain-Specific Vocabulary Instruction Items*

We wrote four items focusing on domain-specific vocabulary instruction and seven items on examining the morphological components of domain-specific vocabulary words. Each of these items had four answer choices with one correct answer. Like the TKVS, these items were scored dichotomously as correct or incorrect. We used both the KnERDI measure (Knowledge

for Enhancing Reading Development Inventory; Davis et al., 2021) and the Mississippi Early Literacy Professional Development Initiative (Folsom et al., 2017) as a close reference for the creation of the 11 items. These items are decidedly more applied than the TKVS as they present teachers with instructional scenarios and ask them to select the best option for teaching vocabulary (Appendix A).

## **Student Measures**

### ***Domain-Specific Vocabulary Knowledge***

To measure depth of domain-specific vocabulary knowledge, students completed a semantic associations task with both science and social studies domain-specific vocabulary words at the conclusion of the MORE content literacy lessons (Read, 2004). A subset of students received exposure to social studies vocabulary while all students were explicitly taught the third-grade MORE science words. The semantic association task measured students' ability to recognize two out of four words that were related to the target word (Kim et al., 2020). We scored each item dichotomously. To answer each question correctly, students were instructed to circle the two words related to the target word. Those who circled only one word received an incorrect score for the item. For the task assessing domain-specific science vocabulary knowledge, item location estimates from 2-parameter logistic (2PL) item response theory (IRT) models ranged from relatively easy in our sample ( $b = -1.04$ ) to very difficult ( $b = 3.67$ ) with an average location parameter estimate of  $-.27$ . For the social studies assessment, location estimates were slightly less variable with locations ranging between  $-.74$  and  $1.07$ , with an average location parameter estimate of  $-.03$ .

### ***Reading Comprehension***

We used two different types of reading comprehension assessments: researcher-developed measures of domain-specific reading comprehension and standardized domain-general reading comprehension assessments. For the researcher developed measures, students completed measures of domain-specific reading comprehension transfer (Kim et al., 2023). By transfer, we refer to students' ability to read and comprehend grade-level passages that covered unfamiliar topics. Using Barnett and Ceci's (2022) taxonomy of transfer, the transfer measures varied both the content (inclusion of taught vocabulary words) and context (passage topics and the extent to which these topics differed from that of the instructional context of the intervention) of the passages along a continuum from near to far. Students first read a near-transfer passage on monkey hearts, a topic closely aligned to the MORE third-grade human body science unit (context). This passage included six of the domain-specific vocabulary words taught in the lessons (content). Next, students read a mid-transfer passage on how scientists worked as a system to send rovers to Mars. The topic of this passage was noticeably more removed from that of the MORE unit on the human body. Thus, the context shifted, although the thread of systems working together remained. We included a few of the domain-specific vocabulary words covered in the lessons, thereby keeping some of the content (i.e., vocabulary) consistent. Finally, the far-transfer passage focused on a much different context – the exploration of the remains of the Titanic. There were no taught domain-specific vocabulary words from the MORE lessons present in this passage (content). Thus, it was a true measure of students' ability to transfer the schema they had instantiated via the MORE lessons (human body systems functioning) to a new but similar topic schema (a system of scientists functioning together).

We calibrated each subtest separately for near-, mid-, and far-transfer assessments. Results from 2PL IRT showed near-transfer location parameters ranging from  $-.52$  to  $.61$

(average  $b = -.25$ ), mid-transfer location parameters ranging from  $-.69$  to  $2.90$  (average  $b = .23$ ), and far-transfer location parameters ranging from  $-.93$  to  $.62$  (average  $b = -.15$ ). Almost all items provided substantial information.

Students took two different standardized assessments in the fall and spring of the academic year. The MAP assessment is a standardized, vertically scaled assessment that the district used to measure academic achievement (Northwest Evaluation Association, 2011). Because the MAP assessment is standardized and not tied to instructional and curricular content, we classified this as a domain-general measure. We included fall scores in our models as a covariate representing baseline achievement for both reading and math. In addition to taking the MAP assessments, all third-grade students within the district completed the state-issued beginning-of-year (BOG) assessment at the beginning of the year and the end-of-grade (EOG) assessment at the conclusion of the school year. Both assessments have documented robust psychometric properties, with the EOG having a Cronbach's alpha reliability estimate of  $.90$  (North Carolina Department of Public Instruction, 2020).

## **Participants**

### ***Teachers***

Participants were 400 third-grade teachers across 109 schools from a large, urban school district in the Southeastern United States. All teachers had previously participated in the MORE intervention as the district had recently adopted MORE as part of its third-grade spring science curriculum. At the conclusion of teaching MORE, teachers participated in a post-intervention survey, of which one section included the TKVS and domain-specific vocabulary knowledge items. We compensated teachers with a gift card for completed the survey.

Participating teachers had a range of experience as classroom teachers including first-year teachers as well as those with 37 years of experience. On average, teachers in the sample had 11.16 years of teaching experience (Standard deviation = 8.14). This Southeastern state had recently passed legislation requiring that all pre-K to third-grade teachers participate in the LETRS (Language Essential for Teachers of Reading and Spelling) training, which focuses on the fundamental components of reading and writing instruction.

### ***Students***

Our sample consisted of 6,977 third-grade students spread across 400 classrooms and 109 schools. On average, there were 17.2 students per classroom. Our sample draws from a diverse school district with over one-third of students coming from high-poverty backgrounds. The majority of students were Black (34%), with Hispanic students constituting the next largest racial group (29%). Nearly a quarter of students were classified as English Language learners (21%). All students within the sample were expected to have participated in the district-adopted MORE third-grade science curriculum, which had been implemented a month prior to teachers completing the end-of-MORE survey. Additionally, students also completed standardized MAP reading and math assessments as well as the BOG and EOG assessments at the beginning and end of third grade.

### **Analysis**

#### ***Research Question 1: Examining the Psychometric Properties of the TKVS Items***

To address our first research question, we began by replicating Duguay and colleagues' (2016) analyses of TKVS items using a Rasch model to estimate item location ( $b$ ) parameters. The Rasch model fixes the item information parameter ( $a$ ) at one, while allowing the item location parameter to vary, thereby providing estimates of each item's difficulty along the theta

scale. The location parameter represents the theta value at which respondents have a 50% probability of answering the question correctly.

Because our sample was comprised of different teachers than those surveyed in the 2016 study, we used a mean-sigma linking procedure to determine if our sample had higher or lower proficiency compared to the original sample of 50 teachers (Embretson & Reise, 2000; Marco, 1977). We estimated the following linking function:

$$\hat{m} = \frac{S_{b_{2022}}}{S_{b_{2012}}}$$

$$\hat{k} = \overline{b_{2022}} - m\overline{b_{2012}}$$

where  $\hat{m}$  tells us how the standard deviation of  $b$  parameters in 2022 differs from the standard deviation of  $b$  parameters in 2012, and how  $\hat{k}$  indicates how an average teacher in 2022 scored on the 2012 scale by determining the difference between the 2022 average ( $\overline{b_{2022}}$ ) and the product of  $\hat{m}$  and the 2012 average ( $\overline{b_{2012}}$ ). Put differently,  $\hat{k}$  allows us to estimate the difference in TKVS scores between 2022 sample and 2012 sample on a common scale. However, the linking results should be interpreted cautiously as we slightly altered TKVS items by eliminating the option, *I don't know*. Limitations of interpreting this analysis are discussed in the results section.

***Research Question 2: Examining the Psychometric Properties of All Items Measuring Teachers' Knowledge of Vocabulary Instruction***

Next, we fit a 2PL IRT model to examine both the discrimination and location parameters of items from the TKVS as well as the 11 additional domain-specific items. The relatively small sample size from the original study included only 50 teachers, likely preventing researchers from fitting a 2PL IRT model. Rasch models assume that all items provide equal information while an advantage of the 2PL IRT model is that the model allows for both the item location and item

discrimination parameters to vary, thus enabling examination of not only item difficulty but also the most and least informative items. Because our sample was substantially larger, we were able to fit a 2PL IRT model. We then estimated Expected A Posteriori (EAP) latent trait scores for each teacher using the *predict* function in Stata 17 for use in subsequent regression models (Chapman, 2022; Muraki & Engelhard, Jr., 1985).

***Research Question 3: Correlating Teachers’ Knowledge of Vocabulary Instruction with Student Outcomes***

To address if and to what extent elementary school teachers’ knowledge of teaching vocabulary correlates with student outcomes, we fit a series of 3-level hierarchical linear models (HLMs) with school random effects at level 3 and teacher random effects at level 2 (Raudenbush & Bryk, 2002). The reduced-form models are specified as follows:

$$Y_{ijk} = \delta_{000} + \gamma_{010}KNOW_{jk} + \sum_{q=2}^3 \gamma_{0q0} L2COV_{qjk} + \sum_{p=1}^{12} \beta_{00p} L1COV_{pijk} + r_{00k} + u_{0jk} + \varepsilon_{ijk}$$

$$r_{00k} \sim N(0, \sigma_r^2)$$

$$u_{0jk} \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$$

where  $Y_{ijk}$  represents the various assessment outcomes (science and social studies vocabulary knowledge, near transfer, mid transfer, far transfer, MAP-Reading, and EOG) for student  $i$  in classroom  $j$  in school  $k$ .  $\delta_{000}$  is the average school intercept.  $\gamma_{010}$  represents the coefficient on teacher knowledge of vocabulary instruction score in standard deviation units and  $\gamma_{0q0}$  ( $q=2,3$ ) are coefficients for two teacher-level covariates (years of teaching experience, past experience teaching MORE intervention).  $\beta_{00p}$  ( $p=1,\dots,12$ ) are coefficients for 12 student-level covariates (race/ethnicity, measure of neighborhood poverty, gender, English Language Learner status, three

beginning-of-the-year test scores).  $r_{00k}$  are the school-level random intercepts,  $u_{0jk}$  are the classroom-level random intercepts, and  $\varepsilon_{ijk}$  are the student-level residuals.

### ***Missing Data***

We excluded 20 participating teachers who did not attempt any items from our analyses. A handful of teachers in the final sample answered only some items, resulting in some incomplete responses. As such, we used EAP scores rather than sum or mean scores to allow for greater precision in our modeling as EAP scores weight items based on each item's information parameter rather than weighting all items equally and adjust for missing data using empirical Bayes procedures, thereby maximizing the information contained in the score (Chapman, 2022; Muraki & Engelhard, Jr., 1985).

## **Results**

All descriptive statistics, psychometric analyses, and regression analyses were conducted using Stata 17 (StataCorp, 2021). Teacher-level descriptive statistics as well as their average scores on each of the domains of the knowledge measure are shown in Table 1. The average teacher had around 11 years of teaching experience and had previously taught roughly one MORE unit in previous years. Teachers correctly answered 75 percent of TKVS items, 72 percent of domain-specific vocabulary items, and 74 percent of all items. As shown in Table 2, teachers' knowledge of vocabulary instruction was weakly correlated with teachers' years of classroom experience ( $r = .25$ ). Both the TKVS and domain-specific items were moderately correlated with each other ( $r = .41$ ), and both sections were strongly correlated with the overall unitary construct, knowledge of vocabulary instruction (TKVS  $r = .89$ , DS  $r = .78$ ).

Table 3 lists the average scores for all items, including the scores from the original study (Duguay et al., 2016) along with scores from the present study with the change in scores between

both samples listed in the far right column where applicable. Our sample of teachers did fairly well and on average scored better on most items compared to Duguay and colleagues' sample.

Students' demographic information and baseline and summative assessment data are shown in Table 4. The sample spanned the continuum of low to high poverty levels and was quite diverse. MAP, BOG, and EOG assessments are listed as scaled scores as we did not have access to item level data. Pairwise correlations for student-level variables are listed in Table 5. Our researcher developed measures (science/social studies vocab assessments, near-, mid-, and far-transfer assessments) were significantly and moderately correlated with all standardized baseline and summative measures (BOG, EOG, MAP Reading & Math Fall/Spring).

### **Research Question 1: Psychometric Properties of the TKVS**

To evaluate the psychometric properties of the TKVS items, we replicated Duguay and colleagues' (2016) analytic procedure by fitting a Rasch model. Results are displayed in Table 6 with the original Rasch model results listed for TKVS items of the original sample along with the conceptual replication Rasch model results of our sample in the column directly to the right. We use the term *conceptual replication* because we did not include all items across all six sections of the TKVS. One benefit of IRT is that parameters should not be sensitive to adding or dropping items if the assumption of local independence is met. Thus, our decision to include some items and exclude others should not impact the parameter estimates. After consulting with psychometricians, we decided to drop the *I don't know* answer option and force respondents to choose *True* or *False*. While all TKVS items function as a unitary construct, we identify the three sections of the survey and their corresponding items for ease of interpretation.

Rasch model results reveal that for the current study sample, on average, items were considerably easier compared to the original study. For example, item 5 in the Development of

Vocabulary Knowledge section had positive difficulty for respondents in the original study ( $b = 1.19$ ) compared to respondents in our sample where the item had negative difficulty ( $b = -.95$ ). More broadly, for the *development of vocabulary knowledge* section, seven of the twelve items in the original study had positive difficulty parameters (items 1a, 3e, 4f, 5g, 6i, 7j, 11n), whereas our analysis revealed only two items with positive difficulty parameters (items 1a, 4f).

For *teaching individual words* items, our analysis indicated that outside of item 3c, all other items had below-average difficulty whereas in the original study, items 1a, 2b, and 3c had above-average difficulty. For items in the *fostering word consciousness* section, Duguay and colleagues (2016) reported that five of the six items had above average difficulty. Yet, our analyses reveal the opposite – that five of the six items had below average difficulty. In sum, the items we selected from the TKVS were relatively easy for third-grade teachers.

We decided to employ the linking function because we had two sets of responses to the same items. As such, we found that compared to respondents in the 2015 study, our sample of respondents scored 1.86 standard deviation units higher, implying that they had greater knowledge of vocabulary instruction. Results from the linking function, however, should be interpreted cautiously. While the items were the same, the answer choices were slightly altered given that we eliminated the *I don't know* option and forced teachers to pick *True* or *False*. By restricting the answer choices to two options, the likelihood of guessing correctly increased, which may inflate correct response rates and decrease item difficulty parameter estimates. Yet, 2015 and 2022 location parameter estimates were positive and strongly correlated ( $r = .84, p < .001$ ), thereby supporting the fact that items from 2015 and 2022 functioned similarly. That is, the relative properties of these items are similar in that the hardest items from 2015 tended to be the hardest items from 2022.

## **Research Question 2: Psychometric Properties of the TKVS and Domain-Specific Vocabulary Items Using 2PL IRT Models**

We examined items from the TKVS measure (Duguay et al., 2016) as well as the 11 items focusing on knowledge of teaching domain-specific vocabulary and examined the psychometric properties using a 2PL IRT model to estimate both the location and discrimination parameters. Results are displayed in two righthand columns of Table 6. Of the 26 TKVS items, three items had negative discrimination parameters and six items had discrimination parameters indistinguishable from zero. Of the 11 domain-specific items, one item had a negative discrimination parameter, and three items had discrimination parameters indistinguishable from zero. As such, we dropped these 13 items for subsequent HLM analyses.

The 2PL IRT model revealed that item information varied substantially, with the more challenging items having low item information (e.g., *dgvocabdevelop1a*, *dgvocabdevelop4f*, *dgvocabteach3c*) and easier items often providing greater information (e.g., *dgvocabdevelop8k*, *dgvocabdevelop9l*, *dgvocabteach2b*, *dgvocabteach6f*, *dgvocabteach7g*, *dgvocabfoster4d*). Thus, the *a* and *b* parameters were negatively correlated ( $r = -.27$ ). Had we solely examined Rasch model results, we would have retained all items, whereas the 2PL model revealed that when both item information and location parameters were allowed to vary, some items functioned the opposite as intended with negatively sloped logistic curves (e.g., *dgvocabdevelop2d*). Forcing items to have uniform item information parameters equal to 1 resulted in misleading information about the item location and information parameters for 13 items. The Cronbach's alpha reliability estimate for the measure was  $\alpha = .73$ .

## **Research Question 3: Correlating Teachers' Knowledge of Vocabulary Instruction with Student Outcomes**

To answer our final research question to determine if and to what extent items from the TKVS and DS combined correlated with student outcomes, we fit a series of HLMs. Results are listed in Table 7. There were no significant associations between teachers' knowledge of vocabulary instruction and student achievement on vocabulary and end-of-year reading comprehension assessments.

### **Discussion**

The findings from this conceptual replication evaluate a revised and expanded measure of teachers' knowledge of teaching vocabulary, which includes items from the TKVS (Duguay et al., 2016) as well as new items focusing on domain-specific vocabulary instruction. Where the originally TKVS study found, using a Rasch model, that item-difficulty varied across a continuum from easy to hard, our analysis using a Rasch model found that many of the TKVS items were actually quite easy. However, we expanded on the original analysis by using a 2PL IRT model. Results indicated that indeed items were still easy, but additionally, not all items provided equal information (the Rasch model fixes all items to have equal information whereas the 2PL model allows the slopes to vary). One question arises: why did the TKVS item-difficulty parameters from our analysis differ drastically from that of Duguay and colleagues (2016)? One possible explanation is that eliminating the *I don't know* answer choice (originally scored as incorrect) increased the likelihood that teachers guessed correctly as teachers were forced to choose *True* or *False*. Yet, the correlation of difficulty parameters between the original study and the present study was .84, suggesting that while we slightly altered the answer choices, the items from both administrations functioned similarly.

Another likely explanation lies in the possible effects of statewide legislation focusing on the science of reading. At the time of the present study, 32 states and the District of Columbia

had adopted laws requiring that teacher-preparation programs and professional development address components of the science of reading (Schwartz, 2023). Teachers in our sample had completed the state-mandated LETRS training, a requirement of the state's science of reading legislation. Perhaps our sample of teachers had greater knowledge than teachers from the original TKVS study (Duguay et al., 2016) of the importance of vocabulary knowledge and instruction and how it supports literacy development. More specifically, given that the majority of TKVS items had negative difficulty parameters in our sample raises the possibility that recent legislation may have increased the baseline level of teacher knowledge of teaching vocabulary, although additional evidence is needed to establish a causal link. Furthermore, the weak correlation between experience as a classroom teacher and knowledge of teaching vocabulary ( $r = .25$ ) reveals that knowledge is largely a function of something other than experience teaching and that standardized coursework on foundational knowledge of literacy practices could explain the increased knowledge of teachers. In fact, the correlation between experience teaching and only TKVS items was even weaker ( $r = .15$ ).

In addition to including items from three sections of the TKVS, our measure also included 11 items focusing on teaching domain-specific vocabulary. We found that like the TKVS items, the domain-specific items were still relatively easy for teachers. However, it is likely that these items were sensitive to practices covered in MORE professional learning as well as the various lessons. That is, MORE teaches domain-specific vocabulary words and has targeted lessons that examine the sublexical aspects of those words. The 11 items could have evoked teachers' memory of the lessons and professional learning as some items included scenarios resembling MORE lessons, thereby possibly making the correct answers that much more obvious to teachers.

Next, we wanted to determine if this altered and expanded measure correlated with student outcomes. One limitation of various measures of teacher knowledge is that these measures' predictive ability on student outcomes is often not examined (Davis et al., 2021; Duguay et al., 2015; McMahan et al., 2019). Unsurprisingly, there were no significant associations between teachers' knowledge of vocabulary instruction and student vocabulary or reading comprehension outcomes. At the more granular level, we did not find evidence of significant associations between the TKVS items and student outcomes or between the domain-specific vocabulary items and student outcomes. One explanation is that because teacher knowledge of vocabulary instruction has improved, the predictive ability of the measure has been attenuated due to limited variation of our sample's knowledge. Alternatively, while results showed that teachers may possess greater knowledge of teaching vocabulary in 2022 compared to nearly a decade earlier – a potential result of participating in state-mandated science of reading professional learning – simply possessing knowledge does not necessarily translate to effective practice. In fact, in the final report for the Reading First study (Gamse et al., 2008), the average instructional time second-grade teachers spent on phonics instruction was nearly three times that of vocabulary instruction. In an analysis of third-grade teachers' support for helping students with vocabulary learning, researchers found that teachers did not give students opportunities to engage with word meanings (Carlisle et al., 2013), and during vocabulary-focused lessons, teachers only engaged students in explaining word meanings 17 percent of the time. Even more surprising, however, was that teachers asked students to explain word meanings “in only 3% of the reading comprehension lessons” (p. 1382). How can students hope to develop strong literacy skills without adequate vocabulary instruction? While TKVS and domain-specific vocabulary items could be used as a diagnostic tool, they should be used in tandem with measurement of

how teachers actually introduce vocabulary words and how they create opportunities for students to engage in using them.

### **Limitations and Future Research**

While our findings indicate that items from the TKVS were easy for our sample of third-grade teachers, we did not include all items from the original measure due to time constraints. It is possible that excluded items from the remaining three sections of the TKVS are challenging – or easier. Second, we deviated from the original instructions in that we restricted answer choices to *True* or *False* and omitted the *I don't know* option. Forcing teachers to choose between two choices instead of three increased the probability of guessing the correct answer when respondents might have chosen *I don't know* in the original administration of the TKVS. Our domain-specific items were also relatively easy, although this may have been confounded by the fact that all surveyed teachers had participated in the MORE intervention, which places significant emphasis on teaching domain-specific vocabulary and morphology. Future research should deploy these items in a state that has not yet passed science of reading legislation or participated in the MORE intervention.

Further research evaluating the measure should be done in conjunction with examining the extent to which teacher knowledge of vocabulary instruction translates to classroom practice (McCutchen et al., 2002) to better understand the relationship between teacher content knowledge and pedagogical content knowledge. While teachers need a strong understanding of the content, they also need adequate *pedagogical content knowledge* – the knowledge that allows them to convey lesson content to learners effectively (Shulman, 1986). One might deduce that as teacher's content and pedagogical knowledge increase, so does the potential for greater effectiveness of conveying content to students. Using audio/video recordings of the classroom

can provide insight into vocabulary instruction as well as the extent to which teachers create opportunities for students to engage in using target vocabulary in discussion.

### **Implications**

The primary goal of this study was to conduct a conceptual replication and extend an existing measure by adding items focusing on domain-specific vocabulary knowledge, a topic of emerging interest since the publication of the TKVS study (Duguay et al., 2016). As such, further replication and expansion of the present study is warranted. Is our sample of teachers more knowledgeable than others? Are there items not yet written that could be more challenging as the majority of items across the current measure were relatively easy for teachers. These questions suggest that there is a need for further replication and expansion. Second, this measure has the potential to be used as a diagnostic tool to assess teachers' knowledge of vocabulary instruction, and this diagnostic information could ultimately be used to help improve instructional quality by aligning it with practices grounded in the science of reading. Third, measuring teacher knowledge helps align instructional practice with policy goals aligned with the science of reading, which can help track improvements in professional learning and its impact on teacher knowledge. Lastly, many states have not yet passed science of reading legislation, leaving open the possibility that some teachers have not yet engaged in professional learning that targets core literacy practices aligned with the science of reading. The TKVS and domain-specific vocabulary items could provide diagnostic information about the knowledge teachers possess about vocabulary instruction and if schools and districts should invest in increasing teachers' knowledge about vocabulary instruction.

### **References**

- Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32(8), 10–14.  
<https://doi.org/10.3102/0013189X032008010>
- Annie E. Casey Foundation. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation*. Baltimore, MD: Author. Retrieved from <https://eric.ed.gov/?id=ED518818>
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). International Reading Association.
- Applegate, A. J., & Applegate, M. D. (2004). The Peter Effect: Reading Habits and Attitudes of Preservice Teachers. *The Reading Teacher*, 57(6), 554–563.
- Bollen, K., Cacioppo, J. T., Kaplan, R. M., Krosnick, J. A., & Olds, J. L. (2015). Social, behavioral, and economic sciences perspectives on robust and reliable science: Report of the Subcommittee on Replicability in Science, Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences. Retrieved from the National Science Foundation Web site: [www.nsf.gov/sbe/AC\\_Materials/SBE\\_Robust\\_and\\_Reliable\\_Research\\_Report.pdf](http://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf)
- Carlisle, J. F., Kelcey, B., Rowan, B., & Phelps, G. (2011). Teachers' knowledge about early reading: Effects on students' gains in reading achievement. *Journal of Research on Educational Effectiveness*, 4(4), 289–321. <https://doi.org/10.1080/19345747.2010.539297>
- Carlisle, J. F., Kelcey, B., & Berebitsky, D. (2013). Teachers' support of students' vocabulary learning during literacy instruction in high poverty elementary schools. *American Educational Research Journal*, 50(6), 1360–1391.  
<https://doi.org/10.3102/0002831213492844>

Chapman, R. Expected a posteriori scoring in PROMIS<sup>®</sup>. *J Patient Rep Outcomes* **6**, 59 (2022).

<https://doi.org/10.1186/s41687-022-00464-9>

Connor, C. M., Dombek, J., Crowe, E. C., Spencer, M., Tighe, E. L., Coffinger, S., Zargar, E.,

Wood, T., & Petscher, Y. (2017). Acquiring science and social studies knowledge in

kindergarten through fourth grade: Conceptualization, design, implementation, and

efficacy testing of content-area literacy instruction (CALI). *Journal of Educational*

*Psychology*, *109*(3), 301–320. <https://doi.org/10.1037/edu0000128>

Connor, C. M., Phillips, B. M., Kim, Y. G., Lonigan, C. J., Kaschak, M. P., Crowe, E., Dombek,

J., & Al Otaiba, S. (2018). Examining the efficacy of targeted component interventions

on language and literacy for third and fourth graders who are at risk of comprehension

difficulties. *Scientific Studies of Reading*, *22*(6), 462–484.

<https://doi.org/10.1080/10888438.2018.1481409>

Davis, D. S., Samuelson, C., Grifenhagen, J., DeIaco, R., & Relyea, J. (2021). Getting KnERDI

with language: Examining teachers' knowledge for enhancing reading development in

code-based and meaning-based domains. *Reading Research Quarterly*, rrq.445.

<https://doi.org/10.1002/rrq.445>

Derksen, M., & Morawski, J. (2022). Kinds of Replication: Examining the Meanings of

“Conceptual Replication” and “Direct Replication.” *Perspectives on Psychological*

*Science*, *17*(5), 1490–1505. <https://doi.org/10.1177/17456916211041116>

Duguay, A., Kenyon, D., Haynes, E., August, D., & Yanosky, T. (2016). Measuring teachers'

knowledge of vocabulary development and instruction. *Reading & Writing*, *29*(2), 321–

347. <https://doi.org/10.1007/s11145-015-9598-4>

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary

instruction on passage-level comprehension of school-age children: A Meta-analysis.

*Journal of Research on Educational Effectiveness*, 2(1), 1–44.

<https://doi.org/10.1080/19345740802539200>

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates Publishers.

*End-of-Grade (EOG) | NC DPI*. (n.d.). Retrieved December 27, 2023, from

<https://www.dpi.nc.gov/districts-schools/testing-and-school-accountability/state-tests/end-grade-eog>

Folsom, J. S., Smith, K. G., Burk, K., & Oakley, N. (2017). *Educator outcomes associated with implementation of Mississippi's K–3 early literacy professional development initiative*. 65.

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First Impact Study final report* (NCEE 2009-4039). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H. S., Doolittle, F., Zhu, P., Szejnberg, L., & Silverberg, M. (2008). The impact of two professional development interventions on early reading instruction and achievement. In *American Institutes for Research, Sep 2008, xxxiii+78p* (p. xxxiii+78p-xxxiii+78p).

Guthrie, J. T., & Klauda, S. L. (2014). Effects of classroom practices on reading comprehension, engagement, and motivations for adolescents. *Reading Research Quarterly*, 49(4), 387–416.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., Scaffidi, N. T., & Tonks, S. (2004). *Increasing Reading Comprehension and Engagement Through*

- Concept-Oriented Reading Instruction. *Journal of Educational Psychology*, 96(3), 403–423. <https://doi.org/10.1037/0022-0663.96.3.403>
- Hall, C., Solari, E. J., Hayes, L., Dahl-Leonard, K., DeCoster, J., Kehoe, K. F., Conner, C. L., Henry, A. R., Demchak, A., Richmond, C. L., & Vargas, I. (2024). Validation of an instrument for assessing elementary-grade educators' knowledge to teach reading. *Reading & Writing*, 37(8), 1955–1974. <https://doi.org/10.1007/s11145-023-10456-w>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kim, J. S., Burkhauser, M. A., Mesite, L. M., Asher, C. A., Relyea, J. E., Fitzgerald, J., & Elmore, J. (2020). Improving reading comprehension, science domain knowledge, and reading engagement through a first-grade content literacy intervention. *Journal of Educational Psychology*. <https://doi.org/10.1037/edu0000465>
- Kim, J. S., Burkhauser, M. A., Relyea, J. E., Gilbert, J. B., Scherer, E., Fitzgerald, J., Mosher, D., & McIntyre, J. (2023). A longitudinal randomized trial of a sustained content literacy intervention from first to second grade: Transfer effects on students' reading comprehension. *Journal of Educational Psychology*, 115(1), 73–98. <https://doi.org/10.1037/edu0000751>
- Kim, J. S., Gilbert, J. B., Relyea, J. E., Rich, P., Scherer, E., Burkhauser, M. A., & Tvedt, J. N. (2024). Time to transfer: Long-term effects of a sustained and spiraled content literacy intervention in the elementary grades: Developmental Psychology. *Developmental Psychology*. <https://doi.org/10.1037/dev0001710>

- Kim, J. S., Relyea, J. E., Burkhauser, M. A., Scherer, E., & Rich, P. (2021). Improving elementary grade students' science and social studies vocabulary knowledge depth, reading comprehension, and argumentative writing: A conceptual replication. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-021-09609-6>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling, fourth edition: Vol. Fourth edition*. The Guilford Press.
- Lesnick, J., Goerge, R. M., Smithgall, C., & Gwynne, J. (2010). *Reading on grade level in third grade: How is it related to high school performance and college enrollment? A longitudinal analysis of third-grade students in Chicago in 1996-97 and their educational outcomes*. Chicago, IL: Chapin Hall at the University of Chicago. Retrieved from <https://eric.ed.gov/?id=ED517805>.
- Lexia. (n.d.). Lexia LETRS suite of professional learning: Language essentials for teachers of reading and spelling. (p. 1-16). <https://www.lexialearning.com/resources/guides/letrs-suite-of-professional-learning-brochure>
- Makel, M. C., Plucker, J. A., Reiter-Palmon, R., & Tinio, P. (2014). Creativity Is More Than Novelty: Reconsidering Replication as a Creativity Act. *Psychology of Aesthetics, Creativity, and the Arts*, 8(1), 27–29. <https://doi.org/10.1037/a0035811>
- Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14, 139–160. <https://doi.org/10.1111/j.1745-3984.1977.tb00033.x>
- McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., Quiroga, T., & Gray, A. L. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities*, 35(1), 69–86. <https://doi.org/10.1177/002221940203500106>

- McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior, 15*(1), 3–18. <https://doi.org/10.1080/10862968309547474>
- McMahan, K. M., Oslund, E. L., & Odegard, T. N. (2019). Characterizing the knowledge of educators receiving training in systematic literacy instruction. *Annals of Dyslexia, 69*(1), 21–33. <https://doi.org/10.1007/s11881-018-00174-2>
- Mosher, D. M., Burkhauser, M. A., & Kim, J. S. (2024). Improving second-grade reading comprehension through a sustained content literacy intervention: A mixed-methods study examining the mediating role of domain-specific vocabulary. *Journal of Educational Psychology, 116*(4), 550–568. <https://doi.org/10.1037/edu0000868>
- Muraki, E., & Engelhard, G. (1985). Full-Information Item Factor Analysis: Applications of EAP Scores. *Applied Psychological Measurement, 9*(4), 417-430. <https://doi.org/10.1177/014662168500900411>
- North Carolina Department of Public Instruction (2020). The North Carolina Department of Public Instruction Mathematics 3-8 End of Grade (EOG) NC Math 1 and NC Math 3 End of Course (EOC) Technical Report 2018-2019. Retrieved from <https://www.dpi.nc.gov/media/10219/open>
- Northwest Evaluation Association. (2011). *Technical manual for measures of academic progress and measures of academic progress for primary grades*. Lake Oswego, OR: Author.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T.

- (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425.  
<https://doi.org/10.1126/science.aab2374>
- Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, 11(4), 357–383. <https://doi.org/10.1080/10888430701530730>
- Phillips, B. M., Oliver, F., Tabulda, G., Wood, C., & Funari, C. (2020). Preschool teachers’ language and vocabulary knowledge: Development and predictive associations for a new measure. *Dyslexia (10769242)*, 26(2), 153–172. <https://doi.org/10.1002/dys.1644>
- Piasta, S. B., Connor, C. M., Fishman, B. J., & Morrison, F. J. (2009). Teachers’ Knowledge of Literacy Concepts, Classroom Practices, and Student Reading Growth. *Scientific Studies of Reading*, 13(3), 224–248. <https://doi.org/10.1080/10888430902851364>
- Read, J. (2004). *Plumbing the depths: How should the construct of vocabulary knowledge be defined* (P. Bogaards & B. Laufer-Dvorkin, Eds.; pp. 209–227). John Benjamins Publishing Company.
- Relyea, J. E., Kim, J. S., Rich, P., & Fitzgerald, J. (2024). Effects of tier 1 content literacy intervention on early-grade English learners’ reading and writing: Exploring the mediating roles of domain-specific vocabulary and oral language proficiency. *Journal of Educational Psychology*. <https://doi.org/10.1037/edu0000882>
- Romance, N. R., & Vitale, M. R. (2001). Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications. *International Journal of Science Education*, 23(4), 373–404.  
<https://doi.org/10.1080/09500690116738>
- Schwartz, S. (2022, July 20). Which States Have Passed “Science of Reading” Laws? What’s in Them? *Education Week*. <https://www.edweek.org/teaching-learning/which-states-have-passed-science-of-reading-laws-whats-in-them/2022/07>

- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. JSTOR. <https://doi.org/10.2307/1175860>
- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, 56(1), 72–110. JSTOR. <https://doi.org/10.2307/1170287>
- StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC.
- U.S Department of Education, Institute of Education Sciences, National center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Reading Assessments & Achievement Levels.
- Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). REPLY TO INBAR: Contextual sensitivity helps explain the reproducibility gap between social and cognitive psychology. *Proceedings of the National Academy of Sciences - PNAS*, 113(34), E4935–E4936. <https://doi.org/10.1073/pnas.1609700113>
- Wright, T. S., & Cervetti, G. N. (2017). A Systematic Review of the Research on Vocabulary Instruction That Impacts Text Comprehension. *Reading Research Quarterly*, 52(2), 203–226. <https://doi.org/10.1002/rrq.163>

**Table 1: Teacher-Level Descriptive Statistics**

	N	M (SD)
Years of Teaching Experience	400	11.25 (8.19)
Past Experience Teaching Intervention	400	.92 (.99)
TKVS Score (%)	395	75 (10)
DS Score (%)	400	72 (17)
Overall Teacher's Knowledge of Vocabulary Instruction Score (%)	400	74 (10)

*Note.* TKVS = Teacher Knowledge of Vocabulary Survey; DS = Domain-Specific; Avg = Average scores.

**Table 2: Pairwise Correlations for Teacher-Level Variables**

	Yrs teaching	Past MORE	TKVS Score	DS Score
Yrs of teaching experience	–			
Past intervention experience	.09	–		
TKVS Score	.15*	-.16*	–	
DS Score	.27*	-.07	.41*	–
Overall Score	.25*	-.14*	.89*	.78*

*Note.* Yrs = Years; Past intervention experience is the number of intervention units teachers had previously taught; TKVS = Teacher Knowledge of Vocabulary Survey; DS = Domain Specific  
\*  $p < .05$ .

**Table 3: Item-Scores for Teachers' Knowledge of Vocabulary Instruction**

Section	2012 <i>N</i>	2012 <i>Avg</i>	2022 <i>N</i>	2022 <i>Avg</i>	$\Delta$ '12-'22
TKVS Development					
1a	50	24%	395	24%	–
2d	50	98%	395	99%	1%
3e	50	52%	395	58%	6%
4f	50	30%	395	41%	11%
5g	50	62%	395	70%	8%
6i	50	32%	395	47%	15%
7j	50	32%	395	52%	20%
8k	50	94%	395	93%	-1%
9l	50	94%	395	96%	2%
10m	50	86%	395	82%	-4%
11n	50	76%	395	87%	11%
12o	50	96%	395	98%	2%
TKVS Teaching					
1a	50	70%	395	72%	2%
2b	50	70%	395	85%	15%
3c	50	62%	395	31%	-31%
4d	50	88%	395	95%	7%
5e	50	98%	395	96%	-2%
6f	50	100%	395	98%	-2%
7g	50	98%	395	90%	-8%
8h	50	90%	395	95%	5%
TKVS Fostering					
1a	50	76%	395	94%	18%
2b	50	96%	395	90%	-6%
3c	50	62%	395	69%	7%
4d	50	76%	395	87%	11%
5e	50	72%	395	78%	6%
6f	50	76%	395	25%	-51%
Teaching DS Vocabulary					
v1	–	–	400	65%	–
v2	–	–	400	48%	–
v3	–	–	400	80%	–
v4	–	–	400	83%	–
v5	–	–	400	83%	–
v6	–	–	400	67%	–
v7	–	–	400	80%	–
v8	–	–	400	92%	–
v9	–	–	400	60%	–
v10	–	–	400	71%	–
v11	–	–	400	63%	–

**Table 4: Descriptive Statistics for Student-Level Characteristics**

Variable	<i>N</i>	<i>M (SD)</i>
Low SES (%)	2,508	36
Mid SES (%)	2,249	32
High SES (%)	2,174	31
Black (%)	2,356	34
White (%)	1,820	26
Hispanic (%)	1,997	29
Asian (%)	550	8
Other (%)	254	4
ELL (%)	1,485	21
Male (%)	3,582	51
Sci Vocabulary (%)	5,627	60 (25)
SS Vocabulary (%)	5,563	53 (26)
Near-Transfer Comp (%)	6,006	57 (28)
Mid-Transfer Comp (%)	5,858	50 (24)
Far-Transfer Comp (%)	5,740	54 (28)
MAP Reading Fall	6,802	185.54 (18.46)
MAP Reading Spring	6,862	195.15 (19.03)
MAP Math Fall	6,808	186.51 (15.72)
BOG	6,764	530.91 (8.81)
EOG	6,892	537.96 (10.22)

*Note.* Sci = Science; SS = Social Studies; Comp = Comprehension; MAP = Measure of Academic Progress; BOG = Beginning-of-Grade assessment; EOG = End-of-Grade assessment.

**Table 5: Pairwise Correlations for Student-Level Variables**

	Sci Vocab	SS Vocab	Near Comp	Mid Comp	Far Comp	MAP R- Fall	MAP R- Spring	MAP M-Fall	BOG
Sci Vocab	–								
SS Vocab	.77*	–							
Near Comp	.65*	.67*	–						
Mid Comp	.62*	.66*	.70*	–					
Far Comp	.65*	.69*	.72*	.71*	–				
MAP R-Fall	.70*	.70*	.68*	.66*	.69*	–			
MAP R-Spring	.72*	.70*	.69*	.66*	.69*	.89*	–		
MAP M-Fall	.67*	.65*	.63*	.60*	.63*	.82*	.82*	–	
BOG	.58*	.62*	.62*	.60*	.63*	.77*	.70*	.66*	–
EOG	.68*	.70*	.70*	.66*	.70*	.85*	.86*	.77*	.75*

Note. Sci = Science; SS = Social Studies; Near Comp = Near Transfer Reading Comprehension; Mid Comp Mid Transfer Reading Comprehension; Far Comp = Far Transfer Reading Comprehension; MAP R = Measure of Academic Progress Reading Assessment; MAP M = Measure of Academic Progress Math Assessment; BOG = Beginning of Year Assessment; EOG = End of Year Assessment.

\*  $p < .05$ .

**Table 6: IRT Results for Knowledge Measure Items**

Indicator Sub-Scale	Item	Original Rasch	Conceptual Replication:	Conceptual Replication: 2PL IRT Model	
		Model Results	Rasch Model	Discrimination	Difficulty
		Item difficulty	Item difficulty		
TKVS: Development of Vocabulary knowledge	dgvocabdevelop1a	3.00***	1.33***	.45**	2.73**
	<b>dgvocabdevelop2d</b>	<b>-2.38*</b>	<b>-4.67***</b>	<b>-.83</b>	<b>5.65</b>
	dgvocabdevelop3e	1.64***	-.39**	.76***	-.50**
	dgvocabdevelop4f	2.67***	.40**	.64***	.61**
	dgvocabdevelop5g	1.19***	-.95***	1.03***	-.98***
	<b>dgvocabdevelop6i</b>	<b>2.56***</b>	<b>.16</b>	<b>-.24<sup>†</sup></b>	<b>-.58</b>
	dgvocabdevelop7j	2.56***	-.10	.58***	-.16
	dgvocabdevelop8k	-1.22*	-2.84***	1.03***	-2.91***
	dgvocabdevelop9l	-1.22*	-3.60***	1.04**	-3.65***
	<b>dgvocabdevelop10m</b>	<b>-.25</b>	<b>-1.71***</b>	<b>.15</b>	<b>-10.23</b>
	<b>dgvocabdevelop11n</b>	<b>.46</b>	<b>-2.16***</b>	<b>.14</b>	<b>-14.02</b>
	<b>dgvocabdevelop12o</b>	<b>-1.65*</b>	<b>-4.19***</b>	<b>.58</b>	<b>-6.96</b>
TKVS: Teaching individual words	dgvocabteach1a	.79*	-1.05***	1.12***	-1.02***
	dgvocabteach2b	.79*	-1.97***	1.34***	-1.71***
	dgvocabteach3c	1.19***	.90***	.28*	2.86*
	<b>dgvocabteach4d</b>	<b>-.43</b>	<b>-3.33***</b>	<b>.49<sup>†</sup></b>	<b>-6.46<sup>†</sup></b>
	dgvocabteach5e	-2.38*	-3.53***	.72*	-4.80**
	dgvocabteach6f	-3.60*	-4.07***	1.30**	-3.51***
	dgvocabteach7g	-2.38*	-2.49***	3.49***	-1.47***
	dgvocabteach8h	-.65	-3.27***	.94**	-3.59***
TKVS: Fostering word consciousness	<b>dgvocabfoster1a</b>	<b>.46</b>	<b>-3.01***</b>	<b>.21</b>	<b>-12.91</b>
	<b>dgvocabfoster2b</b>	<b>-1.65*</b>	<b>-2.40***</b>	<b>-.11</b>	<b>19.41</b>
	dgvocabfoster3c	1.19***	-.91***	1.77***	-.68***
	dgvocabfoster4d	.46	-2.16***	2.70***	-1.36***
	dgvocabfoster5e	.69*	-1.41***	1.19***	-1.32***
	<b>dgvocabfoster6f</b>	<b>.46</b>	<b>1.25***</b>	<b>-1.77***</b>	<b>-.93***</b>

**Table 6 continued**

Indicator Sub-scale	Item	Original Rasch	Conceptual Replication:	Conceptual Replication: 2PL IRT Model	
		Model Results	Rasch Model	Discrimination	Difficulty
		Item difficulty	Item difficulty		
	dsvocab1	–	-.70***	.31*	-1.95*
	<b>dsvocab2</b>	–	<b>.09</b>	<b>.15</b>	<b>.52</b>
	<b>dsvocab3</b>	–	<b>-1.62***</b>	<b>.29†</b>	<b>-4.99†</b>
	dsvocab4	–	-1.85***	.77***	-2.32***
	dsmoprholgy1	–	-1.85***	1.15***	-1.73***
	<b>dsmoprholgy2</b>	–	<b>-.82***</b>	<b>.06</b>	<b>-12.06</b>
	dsmoprholgy3	–	-1.62***	.70***	-2.21***
	dsmoprholgy4	–	-2.80***	1.69***	-2.06***
Teaching Domain- Specific Vocabulary	dsmoprholgy5	–	-.46***	1.90***	-.33***
	<b>dsmoprholgy6</b>	–	<b>-1.06***</b>	<b>-.17</b>	<b>5.38</b>
	dsmoprholgy7	–	-.62***	2.51***	-.40***

Note. TKVS = Teach Knowledge of Vocabulary Survey

†  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

**Table 7: Adjusted Correlations Between Student Outcomes and Teacher Knowledge of Vocabulary Instruction**

	Estimate (SE)						
	Sci Vocab	SS Vocab	Near Comp	Mid Comp	Far Comp	MAP-R	EOG
Intercept ( $\delta_{000}$ )	-.11 (.12)	.00 (.12)	.04 (.12)	-.08 (.11)	-.07 (.11)	.12 <sup>†</sup> (.07)	.00 (.08)
TK of Vocab ( $\gamma_{010}$ )	.00 (.02)	-.01 (.01)	-.03 <sup>†</sup> (.02)	.00 (.02)	.00 (.02)	.01 (.01)	.01 (.01)
Student-Level Covariates							
BOG ( $\beta_{001}$ )	.09 (.01)	.16*** (.01)	.22*** (.01)	.18*** (.01)	.21*** (.01)	.00 (.01)	.20*** (.01)
MAP R-F ( $\beta_{002}$ )	.37*** (.02)	.35*** (.02)	.32*** (.02)	.38*** (.02)	.37*** (.02)	.61*** (.01)	.53*** (.01)
MAP M-F ( $\beta_{003}$ )	.25*** (.02)	.18*** (.02)	.17*** (.02)	.16*** (.02)	.12*** (.02)	.29*** (.01)	.17*** (.01)
Black ( $\beta_{004}$ )	-.10* (.05)	-.15** (.05)	-.06 (.05)	-.12* (.05)	-.14** (.05)	-.03 (.03)	-.04 (.03)
White ( $\beta_{005}$ )	-.04 (.05)	.00 (.05)	.04 (.05)	.01 (.05)	-.03 (.05)	.01 (.03)	.04 (.03)
Hispanic ( $\beta_{006}$ )	-.17** (.05)	-.17*** (.05)	-.01 (.05)	-.05 (.05)	-.06 (.05)	-.05 (.03)	-.02 (.04)
Asian ( $\beta_{007}$ )	-.07 (.05)	-.05 (.05)	.04 (.06)	.02 (.05)	.05 (.05)	-.03 (.03)	.05 (.04)
Low SES ( $\beta_{008}$ )	.13 (.11)	-.03 (.11)	-.09 (.11)	.04 (.11)	.01 (.11)	-.06 (.07)	-.03 (.08)
Med SES ( $\beta_{009}$ )	.16 (.11)	.03 (.11)	-.06 (.11)	.07 (.11)	.07 (.11)	-.05 (.07)	-.01 (.08)
High SES ( $\beta_{0010}$ )	.17 (.11)	.06 (.11)	-.01 (.11)	.11 (.11)	.11 (.11)	-.04 (.07)	.03 (.08)
Male ( $\beta_{0011}$ )	.09*** (.02)	.13*** (.02)	-.03 (.02)	.03 (.02)	.05** (.02)	-.07*** (.01)	.02 (.01)
ELL ( $\beta_{0012}$ )	-.12*** (.03)	-.12*** (.03)	.00 (.03)	.02 (.03)	.00 (.03)	-.08*** (.02)	.00 (.02)
Teacher-Level Covariates							
Yrs teaching ( $\gamma_{020}$ )	.01 (.01)	.02 (.01)	.03* (.02)	.04* (.02)	.05** (.02)	.02* (.01)	.01 (.01)
Past intervention ( $\gamma_{030}$ )	-.01 (.02)	.00 (.01)	-.02 (.02)	-.01 (.02)	-.01 (.02)	.00 (.01)	.01 (.01)
Random Effects (SDs)							
Schools $r_{00k}$	.17	.21	.17	.16	.18	.03	.07
Teachers $u_{0jk}$	.18	.16	.19	.25	.21	.09	.05
Residual $\varepsilon_{ijk}$	.64	.63	.66	.63	.64	.43	.49

Note. TK = Teacher Knowledge; BOG = Beginning-of-Year Assessment; MAP = Measure of Academic Progress Assessment; R-F = Fall Reading Score; M-F = Fall Math Score; R-S = Spring Reading Score; SES = Socioeconomic Status; ELL = English Language Learner; Yrs = Years; Past intervention = Number of intervention units previously taught; Sci = science; SS = social studies; Comp = comprehension; EOG = End-of-Grade Assessment; SE = Standard error.

<sup>†</sup>  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

