Monster, P.I.: Validation Evidence for an Assessment of Adolescent Language That Assesses Vocabulary Knowledge, Morphological Knowledge, and Syntactical Awareness

Assessment for Effective Intervention 2022, Vol. 47(2) 89–100 © Hammill Institute on Disabilities 2020 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/1534508420966383 aei.sagepub.com



Amanda P. Goodwin, PhD¹, Yaacov Petscher, PhD², Jamie Tock, PhD², Sara McFadden, BA¹, Dan Reynolds, PhD³, Tess Lantos, PhD¹, and Sara Jones, MEd¹

Abstract

Assessment of language skills for upper elementary and middle schoolers is important due to the strong link between language and reading comprehension. Yet, currently few practical, reliable, valid, and instructionally informative assessments of language exist. This study provides validation evidence for Monster, P.I., which is a gamified, standardized, computeradaptive assessment (CAT) of language for fifth to eighth grade students. Creating Monster, P.I. involved an assessment of the dimensionality of morphology and vocabulary and an assessment of syntax. Results using multiple-group item response theory (IRT) with 3,214 fifth through eighth graders indicated morphology and vocabulary were best assessed via bifactor models and syntax unidimensionally. Therefore, Monster, P.I. provides scores on three component areas of language (multidimensional morphology and vocabulary and unidimensional syntax) with the goal of informing instruction. Validity results also suggest that Monster, P.I. scores show moderate correlations with each other and with standardized reading vocabulary and reading comprehension assessments. Furthermore, hierarchical regression results suggest an important link between Monster, P.I. and standardized reading comprehension, explaining between 56% and 75% of the variance. Such results indicate that Monster, P.I. can provide meaningful understandings of language performance which can guide instruction that can impact reading comprehension performance.

Keywords

language, achievement assessment

Language knowledge has an intricate relationship with reading comprehension. Theories like the Simple View of Reading (Gough & Tumner, 1986) separate reading comprehension into decoding and linguistic comprehension. As students get older, the contribution of decoding decreases and the role of language increases (García & Cain, 2014). This stems from challenging language demands present in middle school academic texts. Simply put, it is easier to learn to sound out the 200,000 words in academic texts (Nagy & Anderson, 1984) than it is to learn their multiple, nuanced meanings and apply those meanings to parsing ideas conveyed via complex syntactical structures (Nagy & Townsend, 2012).

There are many ways to assess language, but few assessments are used at scale (Adlof & Hogan, 2019; Hart et al., 2015).¹ This is because "the presumed multidimensionality of language ability has practical import" (Language and Reading Research Consortium [LARRC], 2015, p. 1948) such that tests with multiple subtests in different language areas are used for the diagnosis of language disorders rather than to inform general instruction. As an alternative, we provide validation information for a language assessment that uses technology to tackle this complexity to provide scores that inform instruction for a broad swath of students.

¹Vanderbilt University's Peabody College, Nashville, TN, USA ²Florida State University, Tallahassee, USA ³John Carroll University, University Heights, OH, USA

Corresponding Author:

Amanda P. Goodwin, Vanderbilt University's Peabody College, 230 Appleton Pl., Nashville, TN 37203, USA. Email: amanda.goodwin@vanderbilt.edu The assessment is more efficient as it is computer-adaptive (CAT), meaning that items are matched with the level of the learner being assessed, leading to fewer items being needed to determine a valid and reliable score. It also uses technology, meaning that it can be administered and scored at scale, and is delivered in an engaging way as a game. Studies such as Clark et al. (2016) and Mitchell and Savill-Smith (2004) suggest gamifying can increase motivation and self-esteem. Below, we explain our language model and the link between language and instruction.

Model of Language

Language has been conceptualized in many ways (see LARRC, 2015 for a discussion). Our language assessment, Monster, P.I., is grounded in theory and research on the relationship between language and reading comprehension. It focuses on three components of language that are particularly important for young adolescents2: morphology, vocabulary, and syntax (Foorman et al., 2015; Kieffer et al., 2016). Specifically, Perfetti and Stafura's (2014) Reading Systems Framework argues that "word-to-text integration processes are central to comprehension" (p. 30) and involve updating "the situation model [in a way] that integrates a word with a text representation" (p. 29) across words, phrases, and sentences. For adolescents, most words are morphologically complex (Nagy & Anderson, 1984). Thus, in regard to integrating word knowledge, we first considered (a) knowledge of the units of meaning that make up words, like prefixes, suffixes, and root words, as well as how they are combined (i.e., morphology); (b) then broader lexical knowledge (vocabulary); and then (c) knowledge of how those words are put together to convey larger meaning (syntax), with syntax defined as knowledge of word order, grammatical rules, and connectives used to combine words to create meaning (Taylor et al., 2012). The emphasis on text links to the language of schooling, which is dense (i.e., embedded clauses, nominalizations) and uses challenging, morphologically complex words (Nagy & Townsend, 2012; Uccelli et al., 2015).

Multiple research studies provide evidence that language is made up of these three language areas. For example, Kieffer et al. (2016) showed for a Grades 3 to 5 sample and a Grade 6 sample that there is a general factor related to overall language skills and specific factors for morphology, vocabulary, and syntax. Similarly, Foorman et al. (2015) showed that the same model had the best fit for fourth to 10th graders. What this means is that performance on different components of language such as vocabulary, morphology, and syntax overlaps such that performance similarly reflects students' language skills. But performance also is unique, reflecting distinct aspects of language that are likely differentially important to making meaning.

This idea of using modeling techniques to use the complexity of a construct like language to provide additional understandings and instructional guidelines is highlighted by Goodwin et al. (2018). Our language model considers construct-relevant versus construct-irrelevant variance when modeling. Related to language, examples of construct-relevant variance would be skills or areas of language that could be instructionally relevant like ability to identify units of meaning (morphology), access word knowledge (vocabulary), or determine the meaning of a combination of words (syntax). In contrast, construct-irrelevant variance would involve performance related to features of the task like the nature of a multiple-choice task. Here, even if such performance explained variance, educational research would not suggest teaching students how to complete multiple-choice items, but rather instructing the content that the multiple-choice item assesses. Another example might be word-specific versus word-general knowledge with the idea that looking across performance on a group of words is more informative to instruction than considering performance on a single word. We use these ideas to guide our theoretical model: the idea that modeling construct-relevant variance can build a nuanced understanding of the construct and provide instructionally relevant scores.

Our language model also builds on work suggesting some key aspects of language themselves are likely multidimensional (Cho & Goodwin, 2017; Goodwin & Cho, 2016; Goodwin et al., 2017, 2018; Kieffer & Lesaux, 2012). Vocabulary is multidimensional: Knowledge of one aspect relates to others such that a student who knows a word's definition is more likely to be able to recognize synonyms or antonyms of the word, yet each assessment of word knowledge also provides unique information about a student's lexicon. Goodwin and Cho (2016) and Kieffer and Lesaux (2012) confirm this: Goodwin and Cho showed performance on multiple-choice, self-report, and related-word production was related but also tapped different aspects of word knowledge, and Kieffer and Lesaux showed models of vocabulary consisted of "three highly related, but distinct dimensions-breadth, contextual sensitivity, and morphological awareness" (p. 347). Morphology is also likely multidimensional, with Goodwin et al. (2017) reporting best fit of a general factor for morphology (i.e., overlap of all morphological knowledge assessments) and seven specific factors representing the different morphological demands in tasks, many of which link to the different ways that morphemes convey information including semantic, syntactic, phonological, and orthographic information. Yet, no dimensionality studies indicate multidimensionality for syntax (Deacon & Kieffer, 2018). For both vocabulary and morphology, research (vocabulary: Pearson et al., 2007; morphology: Goodwin et al., 2017) and theory (Perfetti, 2007) suggest assessing in multidimensional ways can inform supports between the construct and reading comprehension, yet this is not the case with syntax where additional syntax measures do not explain significant additional variance in reading comprehension (Bowey & Patel, 1988; Cain, 2007). The exception is Brimo et al. (2017), but the syntax measure was a listening comprehension measure, drawing into question its relevance. Hence, we conceptualize vocabulary and morphology as multidimensional but, following the literature, theorize syntax as unidimensional.

Many of the complaints related to assessments of aspects of language such as current vocabulary, morphology, and syntax awareness assessments are that they fail to attend to the multidimensional nature of the constructs. For example, according to Pearson et al. (2007), few studies assess vocabulary in a valid manner, partly because current measures rarely consider multiple aspects of word knowledge within assessments. The same concerns are noted for morphology (Bowers et al., 2010; Nagy et al., 2014). Furthermore, assessments that consider the multidimensional nature of vocabulary (Hadley et al., 2019) tend to be individually administered, which is a challenge for classroom operationalization. We developed our work attending to these concerns as scores on different aspects of language likely have different instructional implications.

Written Orthography

In creating our assessment, we took a developmental view. With more formal literacy instruction, it is important to consider how the written orthography conveys meaning, moving beyond a purely oral view of language. For example, *know* and *knowledge* overlap in orthography but not in pronunciation, suggesting that written language conveys important clues to meaning that should be considered when assessing language knowledge for adolescents. In addition, written language tends to include words that are abstract, like *analysis*, which is difficult to assess with pictures. Therefore, our assessment takes into account developmentally the importance of assessing language within the written orthography.

Links to Instruction

The consideration of language as multidimensional is important for instruction because the areas of language and skills tend to be more malleable than general language within these areas. The recipe for instruction—how much vocabulary, morphology, and syntax instruction—is unknown. Research with preK-12 students emphasizes that vocabulary (Elleman et al., 2009; Wright & Cervetti, 2017), morphology (Bowers et al., 2010; Goodwin & Ahn, 2010, 2013), and syntax (Graesser et al., 2011) are teachable, with potential benefits to students with reading difficulties versus typical students (1.23 vs. 0.39 effect size; Elleman et al., 2009). Language data can inform such recipes of instruction.

Current Study

Our study explores the reliability and validity of Monster, P.I. First, we examine the dimensionality of vocabulary, morphology, and syntax with the goal of identifying construct-relevant variance related to teachable skills present in each language component. Second, we explore whether scores are reliable and valid. As part of this, we investigate links between Monster, P.I. and standardized reading vocabulary and reading comprehension.

Method

Participants

Across the 3 years of the study, we worked in an urban district in the Southeastern United States. For our first research question, we used our full sample of 3,214 fifth through eighth graders (n = 1,026 fifth graders, 742 sixth graders, 715 seventh graders, and 731 eighth graders) learning in the classrooms of 15, 38, and 37 teachers in Years 1, 2, and 3, respectively.³ Demographic data from the district for 3,041 participants suggest the sample was 53.4% female and 33.3% Black, 40.4% White, 21.9% Latino, 4.3% Asian, and <1% American Indian. Fifty percent of students were classified as economically disadvantaged, and 8% were identified as English language learners (ELLs). For the second research question, we used a smaller sample that had taken the CAT version of Monster, P.I. during Year 3 along with the standardized measures. This sample included 1,002 students who were 53% female and were 42.8% White, 30.4% Black, 22.3% Latino, and 4.5% Asian. This sample had 30.6% students classified as economically disadvantaged, and 7.3% were identified as ELL.

Development

Year 1 development focused on exploring multidimensionality (i.e., of language and its components). We started with 10 morphological tasks, four vocabulary tasks, and one syntax task.⁴ Tasks and items were either adapted from the literature or created by a panel of experts. Poorly performing tasks due to ceiling effects, poor discrimination, lack of range, and so on were discarded. Results indicated no effect of either reading items aloud (i.e., Cohen's d = -0.23) or order effects (i.e., average Cohen's d = -0.16; range d =-0.37 to 0.08). Year 2 included further piloting and operationalization via the assessment and confirmation of the dimensionality modeled in Year 1. Year 3 included delivery of the CAT, further item piloting, and exploration of links to standardized reading. Years 2 and 3 also included gamification efforts, with Year 2 developing the storyline and worlds and Year 3 including additional 30-s to 1-min games.

Ultimately, Monster, P.I. features a mischievous monster wreaking havoc on a city, including within a school, museum,

and "catch the monster!" At the completion of each area (i.e., a few tasks), participants get a clue and play a 30- to 60-s mini game. The CAT (i.e., using a two-parameter logistic item response adaptive set of algorithms) takes between 20 and 40 min.

Measures

Measures are described below. All tasks were provided in print. Vocabulary and Syntax tasks were also read aloud (and could be repeated). Measures retained for the CAT are starred.

Morphology

*Odd man out (OMO)**. This was adapted from Ku and Anderson (2003). Students identified two of the three words that shared morphemes by clicking on the word that did not fit in the set. Examples include *corner*, *farmer*, and *swimmer* (where the odd man out is *corner*) and *season*, *seashore*, and *seaweed* (answer: *season*). Items varied on whether words overlapped in suffixes, prefixes, or root words and whether relationships were transparent or opaque.

Meaning puzzles (MP)*. This was adapted from intervention work (Goodwin, 2016) and piloting (Pacheco & Goodwin, 2013). Students identified meaning overlap in morphologically complex words by choosing words that shared morphemes. They chose from four answer choices the word that shared a morpheme with the target word. Distractors had visual overlap, but only the correct answer had morphological overlap. Examples include *astronomy*, with choices of *fast*, *strong*, *as*, and *astronaut*, and *divisibility*, with choices of *divide*, *visible*, *sibling*, and *ability*.

Real world suffix (RWS)*. This was adapted from Tyler & Nagy (1989). Students identified the syntactically appropriate form of a root word from four choices to complete a sentence. An example was *It was a ______ occasion when Barack Obama was elected president*. Students had to choose from choices of *historic, historian, history,* and *historically*.

Making it fit (MIF)*. This was adapted from Carlisle (1988). Students read a low-context sentence with a missing word and were given the base form of the missing word. They had to type the form of the word with that base that fit the sentence. For example, *Some people argue that the*

[sense] thing for Rosa Parks to do would have been to give up her seat, but instead, Ms. Parks stood up for what was right and started a movement (Correct: sensible). Word detectives (WD)*. This was adapted from reading vocabulary tasks and Tyler and Nagy (1989). Students identified the meaning of the morphologically complex underlined target word given four choices. To increase motivation, the task was framed as a detective activity, with students encouraged to look for clues in the word and sentence to help figure out the word's meaning. An example was *There was movement in the upper body. The upper body was*: (a) *staying still*, (b) *full of bones, and* (c) *changing position.*

Sentence sense (SS). Students were presented with a morphologically complex word in a sentence. Participants had to apply the meaning to choose the inference that best fit the situation. For example, participants read, *After the hurricane, the city government set out to repair the streets only in residential areas.* The participant then choose the most accurate inference (i.e., apply the meaning of *residential* to the situation) from four choices (e.g., *People needed to get to and from their homes safely. [correct] vs. Major highways had to be rebuilt.*).

Morphological coherence (MC). This was adapted from core academic skills (CALS-I; Uccelli et al., 2015). Participants read a three- to four-sentence passage in which the final sentence had an underlined morphologically complex word that students had to match with earlier text that refers to it. For example: *Complex engineering problems* are *impossible to manage* [correct] unless solved through teamwork. Construction workers and project managers *have to work together*. Often, projects that look <u>unmanage-</u> able can be achieved by people working together.

Morphological spelling (SP)*. Adapted from Carlisle (1988) to assess students' ability to spell morphologically complex words. Students listened using headphones and were asked to spell the word heard using the iPad's keyboard. Examples include *selective* and *odorous*. This task was presented orally and students could repeat the prompt.

Morphological word reading (WR)*. Students listened to three pronunciations of a morphologically complex word and chose the correct pronunciation. Distractors were mispronunciations most often used by middle schoolers based on pilot reading responses. An example is ______, (a) *selective*, (b) *selecteyeve*, and (c) *seelecteyeve*.

Word hunt (WH). This was adapted from Goodwin (2016). Students identified 12 to 14 morphologically complex words from a larger line of words presented on a single line without any spaces between the words (i.e., *camouflageobstaclediagramindigoshovel*). Participants tapped where the word breaks should be.

Vocabulary

Definition (VOD). Students identified the correct meaning of the word via a traditional multiple-choice task. An example is *Which of the following is the correct meaning* for the word "sweeping"? (a) Incomparable in beauty, (b) Having a wide scope or range, and (c) Making fresh.

Antonym/synonym (VOA). Students identified words with the same (synonym) or opposite (antonym) meaning of a given word. An example is Which of the following is the opposite of the word famine? (a) Adequacy, (b) Success, and (c) Plenty. This task tends to be more difficult in item difficulty compared with word definitions (Spencer et al., 2015).

Word relations (VOR). Students had to correctly discern the relationship between two word pairs, one of which contains the target vocabulary word. An example is *Just* is to *unequal* as *smooth* is to: (a) *Rough*, (b) *Surface*, and (c) *Soft*. This task follows Tzuriel and George's (2009) work suggesting that verbal analogy is likely part of a vocabulary construct.

Polysemy (VOP). Students identified multiple meanings of a given target word by answering five questions about each word with two to four meanings being accurate. For example, the target word *charmed* was tested by multiple items such as *Does "charm" mean a piece of a bracelet? Does "charm" mean make mad? Does "charm" mean delight or please?* Students responded yes or no to each question accordingly. This task attended to research suggesting knowing multiple meanings links to reading comprehension (Logan & Kieffer, 2017).

Syntax

Sentence correctives (SYN). This was adapted from Foorman et al. (2017). Students were presented with two or three sentences. They determined the best answer that combined those sentences into one new sentence. Five types of connectives used were additive (i.e., and), causal (i.e., so), temporal (i.e., before), logical (i.e., similarly), and adversative (i.e., but). Distractors contained connectives that were incorrect, changing the overall meaning. An example is *Gabby spent most of the summer at a camp. She liked it. She was eager to return home.* Students had to choose the correct answer: *After spending most of the summer at camp, Gabby was eager to return home, although she liked it.*

Standardized measures. *Reading Comprehension* (Measures of Academic Progress [MAP]; Northwest Evaluation Association [NWEA], 2017) was assessed via the school district using the MAP, which is a multiple-choice CAT aligned to standards and nationally normed. It is designed to assess achievement and growth in reading as well as provide instructional information and link to high-stakes tests. Strong reliability has been shown (.90–.95), as well as

concurrent validity. Overall reading scores derived from reading comprehension performance on literature and informational texts as well as performance on vocabulary items were used.

Vocabulary (Gates–MacGinitie Reading Vocabulary Assessment; MacGinitie et al., 2000) was assessed at the end of the first testing session using Form S of the Level 5 through 8 Gates–MacGinitie Standardized Reading Vocabulary Assessment. Students answered 45 multiplechoice items where they read an underlined word within a phrase and choose the word or phrase that conveyed a similar meaning. This task is used extensively in research with strong reliability and validity such that the Kuder-Richardson (K-R 20) reliability was .90–.92. Extended scale scores were used.

Procedures

Students took the assessment on iPads with headphones. Upon completion, students re-played some of the mini games and built their own monster. Two sessions were used to deliver the items (Years 1 and 2: two sessions of fixed items, Year 3: CAT version in Session 1 and additional fixed items in Session 2). A common-item, nonequivalent research design was used with planned missingness to maximize item deployment while minimizing testing time with individuals. Approximately 20% to 30% of items across samples were common items, whereas the rest were unique items tested. The common items served as anchor items for item-bank building.

Analysis Plan

Multiple-group item response modeling (MG-IRM). To explore our first research question, a series of multiple-group unidimensional and multidimensional IRT models were fit to the item-level data within morphology, vocabulary, and syntax. This served to evaluate item properties (e.g., the item difficulty and discrimination) across a developmental scale of ability (i.e., vertical equating). MG-IRM allowed for simultaneous testing of the factor structure for the items, the vertical equating of item difficulty, and the vertical scaling of person ability. We began by testing whether the tasks were individually unidimensional and then explored skill-level models for morphology and trait-level models for morphology, vocabulary, and syntax. The supplemental online materials detail the skill and trait models for each task.

Reliability. For Research Question 2, the precision of resulting scores for the factors retaining construct-relevant variance was evaluated to understand broad-level reliability. IRT-based reliability differs from classical test theory as the former is reliant on individual-based estimates and the latter is based the assumption of equal reliability given the total test score. Marginal reliability (Sireci et al., 1991) using the

	Grade 5			Grade 6			Grade 7			Grade 8		
Task	n items	М	SD	n items	М	SD	n items	М	SD	n items	М	SD
Combined data	a (N = 3,240)										
Morphology		,										
OMO ^a	28	0.59	0.26	27	0.66	0.30	32	0.67	0.26	28	0.68	0.25
MPU ^a	28	0.48	0.27	33	0.60	0.28	34	0.64	0.29	27	0.60	0.28
MIF ^a	28	0.29	0.30	34	0.41	0.34	33	0.44	0.32	29	0.45	0.31
RWS ^a	36	0.54	0.28	33	0.56	0.31	37	0.63	0.27	29	0.67	0.27
MCO	4	0.66	0.34	2	0.73	0.36	4	0.73	0.32	2	0.76	0.35
SSE	3	0.48	0.32	2	0.49	0.39	I	0.57	0.50	2	0.62	0.38
WH	5	0.69	0.31	7	0.57	0.26	7	0.61	0.25	4	0.56	0.28
WD ^a	32	0.44	0.27	34	0.56	0.27	36	0.58	0.27	29	0.62	0.26
SP ^a	13	0.35	0.30	12	0.37	0.31	12	0.41	0.31	13	0.44	0.30
WR ^a	10	0.71	0.27	8	0.66	0.36	10	0.86	0.26	11	0.79	0.24
Vocabulary												
VOAª	23	0.46	0.32	31	0.52	0.32	32	0.63	0.35	26	0.59	0.29
VOD ^a	27	0.55	0.38	34	0.52	0.32	27	0.57	0.28	22	0.67	0.24
VOR ^a	37	0.39	0.33	52	0.46	0.28	50	0.58	0.29	38	0.51	0.37
VOPa	74	0.68	0.19	81	0.78	0.19	83	0.78	0.19	54	0.78	0.19
Syntax												
SYNª	18	0.52	0.26	21	0.53	0.29	26	0.55	0.28	31	0.59	0.28

Table 1. Descriptive Statistics for Combined Data Over 3 Years.

Note. N items = number of items per task; SYN = syntax; VOA = vocabulary-antonyms/synonyms; VOD = vocabulary-definitions; VOR = vocabularyword relations; VOP = vocabulary-polysemy; OMO = odd man out; MPU = meaning puzzles; MIF = making it fit; RWS = real word suffix; MCO = morphological coherence; SSE = sentence sense; WD = word detectives; SP = spelling; WR = word reading; WH = word hunt. ^aDenotes measure retained in the final CAT.

factor scores (i.e., student ability scores) and standard errors from the MG-IRT was estimated with

$$\bar{\rho} = \frac{\sigma_{\theta}^2 - \sigma_{e^*}^2}{\sigma_{\theta}^2}$$

where σ_{θ}^2 is the variance of ability score for the normative sample and $\sigma_{e^*}^2$ is the mean squared error.

Concurrent, construct, and predictive validity. For Research Question 2, validity was explored via correlations and using hierarchical multiple regression modeling to explore links between performance on the Monster, P.I. assessment and standardized measures of reading achievement. The primary emphasis in Monster, P.I. is on morphology; hence, Step 1 included all morphology skills, Step 2 added vocabulary, and Step 3 added syntax. At each phase of modeling, R^2 was taken so that the incremental value added for each predictor could be evaluated.

Results

Research Question 1: MG-IRM

Because data were collected over a 3-year period with a common-item, nonequivalent group design, data were specifically missing completely at random due to the planned missing data aspect of item deployment across samples. Table 1 includes descriptive statistics, with the number of items administered for each of morphology, vocabulary, and syntax by grade level, the mean percentage correct, and the standard deviation. What is noticeable about the pattern of responses is the developmental increases in the mean percentage grade by grade level for some tasks (OMO, MIF, RWS, MCO, SSE, WD, and Syntax) but not others (MPU, WH, SP, and WR tasks, along with all four vocabulary tasks). Such a phenomenon is expected prior to item culling and validation as creating stable linking items ensures a vertical scale.

Summary of the task-level and skill-level MG-IRMs are reported in online supplemental Table S1 for morphology and Table S2 for vocabulary. Note that the four-factor correlated model of morphology did not converge with the inclusion of higher convergence criteria and increased iterations. Remaining models all demonstrated excellent fit: root mean square error of approximation (RMSEA) point estimates ranged from .000 to .048; the lowest comparative fir index (CFI) observed was .95; and the lowest Tucker– Lewis index (TLI) was .95. For skill-level morphology model results, the bifactor provided the best fit to the data for Skill 1, $\chi^2(117) = 177.65$, CFI = .95, TLI = .94, RMSEA = .031 (90% confidence interval [CI] = [.021, .039]), and Skill 2, $\chi^2(75) = 73.75$, CFI = 1.00, TLI =

Trait/Skill	Task	Skill/Task Name	Marginal Reliability	No. of Final Items	
Morphology					
Skill I		Morphological Awareness	.90	76	
	Task I	Odd man out	.76	41	
	Task 2	Meaning puzzles	.75	35	
Skill 2		Morphological-Syntactic Knowledge	.93	121	
	Task 3	Making it fit	.92	62	
	Task 4	Real word suffix	.85	59	
Skill 3		Morphological-Semantic Knowledge	.70	121	
	Task 5	Word detectives	.70	81	
	Task 6	Morphological coherence		20	
	Task 7	Sentence sense		20	
Skill 4		Morphological-Orthographic/Phonological Knowledge	.92	95	
	Task 8	Word hunt		6 (77)	
	Task 9	Spelling	.87	45	
	Task 10	Word reading	.83	44	
Vocabulary		-	.80		
Skill I	Task 11	Definitions	.40	35	
Skill 2	Task 12	Word relations	.31	39	
Skill 3	Task 13	Verbal analogy	.72	59	
Skill 4	Task 14	Polysemy	.47	24 (102)	
Syntax	Task 15	Connectives	.70	34	

Table 2. Trait-, Skill-, and Task-Level Reliability and Number of Items in CAT Item Bank.

Note. Monster, P.I. computer-adaptive (CAT) presently provides skill-level and trait-level ability scores and not task-level ability scores. Task-level reliability is provided for comprehensive reporting purposes. The number of items for Word Hunt is 6 unique item-bundles that are inclusive of 77 total components. The number of items for Polysemy is 24 unique item-bundles that are inclusive of 102 total components. Marginal Reliability I = estimate reliability from item-level analysis with N = 3,240; Marginal Reliability 2 = estimated reliability from Year 3 CAT trial with N = 1,450.

1.00, RMSEA = .000 (90% CI = [.000, .023]). For both Skills 3 and 4, each of the unidimensional, correlated trait, and bifactor models yielded acceptable fit with CFI and TLI at or above ~.95 and RMSEA <.05.

Result for the trait-level models is provided in Table S1 (morphology) and Table S2 (vocabulary and syntax). For morphology, the bifactor model that included tasks as specific constructs and the four morphology skills as uncorrelated, global constructs fit the data well, $\chi^2(13, 666) =$ 15,250, CFI = .97, TLI = .96, RMSEA = .015 (90% CI = [.012, .025]), as did a trifactor model that included tasklevel factors, skill-level factors, and a global construct of morphology, $\chi^2(13, 524) = 14,445$, CFI = .98 TLI = .98, RMSEA = .011 (90% CI = [.009, .017]). For vocabulary, the bifactor model that included specific factors for each task and a global factor fit the data acceptably, $\gamma^2(375) =$ 476.51, CFI = .93 TLI = .92, RMSEA = .024 (90% CI = [.017, .030]), yet the bifactor model that included wordlevel, specific factors and a global factor provided excellent fit to the data, $\chi^2(228) = 214.39$, CFI = 1.00, TLI = 1.00, RMSEA = .000 (90% CI = [.000, .015]). For syntax, the unidimensional, correlated trait, and bifactor models provided acceptable fit to the data; however, for parsimony purposes, the unidimensional model was selected, $\chi^2(152)$ = 174.89, CFI = .97, TLI = .97, RMSEA = .017 (90% CI = [.000, .028]).

The viability of multiple trait-level models for morphology in the probit estimation (see Table S1 with TLI, CFI, and RMSEA values) led to an important consideration in the estimation of logit-based MG-IRM (see Table S3 for log likelihood, Akaike information criterion [AIC], and Bayesian information criterion [BIC] values)-specifically, the balancing of test information (i.e., reliability) gained by the model with malleability of the factor. In other words, we considered additional information about the operationalization of the model. For example, whereas a trifactor model for morphology would result in a global factor that would theoretically represent the most information about student performance, for the bifactor model, the skilllevel constructs would represent the most reliable portion of the data and yield malleable factors that facilitate the provision of instructional recommendations for teachers. Details of the logit-based MG-IRMs from flexMIRT software are described in supplemental online materials as well as in Table S3 for morphology, vocabulary, and syntax.⁵

A summary of trait-, skill-, and task-level marginal reliability and final item sets are reported in Table 2. Skill 1 represents Morphological Awareness, Skill 2 represents Morphological-Syntactic Knowledge, Skill 3 represents Morphological-Semantic Knowledge, and Skill 4 represents Morphological-Orthographic/Phonological Knowledge. Marginal reliability is presented for the subsample of 1,002

	Variable		В	Bivariate C	orrelations		Multiple Regression R ²			
Grade		Skill I	Skill 2	Skill 3	Skill 4	Voc	Syn	Variable	Y3 Voc	Y3 RC
5	Skill I	I						Morph Skills	.54	.52
	Skill 2	.63	I					Morph + Vocab	.65	.59
	Skill 3	.55	.56	I				Morph + Vocab + Syntax	.65	.62
	Skill 4	.54	.60	.47	I					
	Vocabulary	.51	.58	.49	.41	I				
	Syntax	.49	.42	.46	.47	.40	I			
6	Skill I	I						Morph Skills	.60	.52
	Skill 2	.64	I					Morph + Vocab	.70	.61
	Skill 3	.58	.59	I				Morph + Vocab + Syntax	.70	.63
	Skill 4	.52	.54	.47	I					
	Vocabulary	.55	.57	.52	.40	I				
	Syntax	.40	.44	.48	.44	.46	I			
7	Skill I	I						Morph Skills	.54	.44
	Skill 2	.63	I					Morph + Vocab	.70	.58
	Skill 3	.57	.62	I				Morph + Vocab + Syntax	.71	.59
	Skill 4	.52	.59	.50	I			. ,		
	Vocabulary	.61	.58	.54	.46	I				
	Syntax	.45	.37	.44	.42	.43	I			
8	Skill I	I						Morph Skills	.50	.51
	Skill 2	.69	I					Morph + Vocab	.61	.56
	Skill 3	.62	.66	I				Morph + Vocab + Syntax	.63	.57
	Skill 4	.55	.62	.53	I			. ,		
	Vocabulary	.60	.55	.60	.45	I				
	Syntax	.46	.49	.54	.40	.48	I			

Table 3. Concurrent Correlations Among Monster, P.I. Ability Scores by Grade Level.

Note. Voc = vocabulary, Syn = syntax; Y3 Voc = gates vocabulary; Y3 RC = MAP/Measures of Academic Progress Reading Comprehension Score.

who took a CAT version of Monster, P.I. during Year 3 that included items from Years 1 and 2. Marginal reliability for Skill 1 was estimated at .90 compared with .93 for Skill 2, .70 for Skill 3, and .92 for Skill 4. Marginal reliability for vocabulary was estimated at .80 for the adaptive version of the assessment and .70 for syntax.

Concurrent, Construct, and Predictive Validity

Table 3 reports correlations among the morphology, vocabulary, and syntax scores. Morphology skills were moderately correlated with each other across the grade levels with a lower bound of r = .47 between Skills 3 and 4 in Grades 5 and 6 to an upper bound of r = .69 between Skills 1 and 2 in Grade 8. Vocabulary was moderately correlated with all four morphology skills ranging from r = .40 between vocabulary and Skill 4 in Grade 6 to r = .61 between vocabulary and Skill 1 in Grade 7. Syntax was also moderately correlated with the morphology skill ranging from r = .37 with Skill 2 in Grade 7 to r = .54 with Skill 3 in Grade 8. Vocabulary and syntax correlated in a range from r = .40 to .48 across Grades 6 to 8.

In terms of predictive validity, summary results for hierarchical regression inclusive of the three-stage model building (i.e., all morphology skills, followed by vocabulary and then by syntax) are presented in Table 3 via R^2 for each model by outcome and grade level. Specific results for each model including the beta weights, standard errors, and *p*-values are present in the supplementary materials. Grade 5 results showed that the combination of morphology skill ability scores resulted in 52% to 54% of the variance explained in the selected outcomes compared with 52% to 60% in Grade 6, 44% to 54% in Grade 7, and 50% to 51% in Grade 8. Adding in both vocabulary and syntax indicated 62% to 65% of the variance in Grade 5 outcomes was explained as was 63% to 70% of the variance in Grade 6, 59% to 71% of the variance in Grade 7, and 57% to 63% of the variance in Grade 8.

Discussion

This study provides reliability and validity evidence for a CAT, gamified, standardized assessment of language for students from Grades 5 to 8. What is important about this assessment is that it provides scores at smaller levels (i.e., language areas and/or skills within language areas), which can inform instruction in ways that general language scores do not. It does this while engaging students in answering

questions framed in a gaming environment. Our findings confirm and extend what is present in the literature, making clear advances in understanding how language can be assessed.

Assessment Advancements

Currently, language beyond vocabulary knowledge is rarely assessed in the standardized tests that schools and teachers use to guide their instruction (see Adlof & Hogan, 2019, for a discussion). That is largely due to the complicated nature of the language problem space and the challenges in assessing language as current assessments often take a long time, are individually administered, or do not assess the multidimensional nature of the areas of language that support reading comprehension. As such, teachers do not have language data to guide their instruction, and hence, fine-tuned language instruction in schools is rare.

Our findings indicate that Monster, P.I. can provide valid and reliable scores that can inform instruction—and the assessment capitalized on CAT and used a gamified format—indicating that teachers may be able to ascertain such scores in a more engaging context than traditional testing environments. The scores provided by Monster, P.I. can provide a detailed view of how the student is using their language skills to support reading comprehension, which can inform instruction that takes into account students' strengths and weaknesses.

Building on LARRC's (2015) call to assess "the extent to which the various theorized dimensions of language do indeed represent latent abilities at a given point in a child's development" (p. 1948), our study confirms that for young adolescents, performance on morphological, vocabulary, and syntax together meaningfully represents a student's general language knowledge. In addition, we extend this work to emphasize that morphology and vocabulary themselves are multidimensional such that they are made up of multiple skills or factors that represent construct-relevant variance that is practically meaningful to both researchers and practitioners. In contrast, no evidence for multidimensional syntax was found. This framework provides a foundation for assessments and confirms the importance of attending to multidimensionality in assessing language to inform instruction.

Our assessment also further unravels the complicated and thorough ways that language supports reading comprehension. Our analyses show each skill or component of language had a meaningful relationship to reading comprehension. With that said, considering multiple skills and areas of language highlighted the more extensive role of language in reading comprehension. Therefore, to understand how language is affecting reading comprehension, assessments must provide multidimensional information. From there, scores can drive research-based instruction in

The contribution of language can be highlighted by reflecting on an example. Say a middle schooler encounters Politicization of the indecisive nature of, in a text. Parsing this text requires morphology, vocabulary, and syntax knowledge to build meaning. For example, a middle school reader would consider the units of meaning within politicization and indecisive and connect to their larger morphological families such as political, politically, politicize and decide, decision, decisively, indecisively. The middle school students would also use the syntactic morphological information in the suffixes present in *politicization*, which convey noun meaning the process of becoming political, and in indecisive, which convey adjective describing. That student would also put the meanings of morphemes (i.e., use semantic morphological information) together to figure out these unfamiliar words and then use the phonological/orthographic morphological information to support their morphological word reading skills and building of orthographic representations via morphological spelling skills. An example is that indecisive is made up of the meanings and pronunciations and spellings of in (not) and decisive. At the same time, the reader would consider the definition of indecisive (lack of commitment to a decision) or link to synonyms like ambivalent or hesitant. Then he would put the word information together across the phrase, considering word order and structure of the phrase, linking it to two ideas that in less complex texts might have been presented in two sentences (i.e., syntax, Someone had an indecisive nature. Someone made used that nature in a political way). By providing scores related to each of these skills and language areas, researchers can better understand how language links to reading comprehension and practitioners can better design instruction to meet student's needs. Monster, P.I. can provide teachers with scores that acknowledge the role of language in supporting reading comprehension as both broader and more specific than specified previously.

Limitations

While moving the field forward, there are certain limitations to take into consideration. As mentioned, this is just one conceptualization of language and theory, and research indicates many more areas of language that could be potentially supportive of reading comprehension and give a broader view of the role of language in reading comprehension. In fact, others have argued that considering just morphology, vocabulary, and syntax is too simplistic (Uccelli et al., 2015), but we felt these areas had the most grounding in theory and research. Another challenge relates to the number of items needed to assess these constructs. We built out the morphology section of the assessment, resulting in stable skill-level estimates, but for vocabulary, while a bifactor model fit best, we continued to use the general factor in our modeling due to a combination of less reliability in the word-level factors and the fact that we are keying into the part of the factor structure that is most instructionally relevant. In addition, following guidance from the literature (i.e., no dimensionality studies suggesting multidimensionality for syntax), we used a single syntax task, although considered potential multidimensionality within the task. This task included connectives (Crosson & Lesaux, 2013) as part of combining sentences, so future research should consider a broader range of syntax tasks to confirm dimensionality and future assessment tasks may need to be adjusted based on such findings. In addition, due to limits of testing time, we were unable to compare performance with standardized morphology and syntax measures. We hope future studies will continue to explore these questions. Another limitation involves our focus on students in general rather than individual differences. It is likely that our assessment may differ for groups of readers and we hope future work will unravel this.

Use in Classrooms

This study suggests Monster, P.I. can provide meaningful data for researchers and practitioners. Goodwin et al. (2019) include more information on how Monster, P.I. can inform instruction, but we close here with statements of three teachers from a pilot study who described how Monster, P.I. informed their instruction. Teachers reported that Monster, P.I. "confirmed the major problems that students have struggled with throughout the year"; "helped with intervention instruction to meet each student at their point of need"; and "included more specific information on vocabulary skills." The teachers used this information "to create small group lessons, mini lessons, or flipped lessons for individuals"; "to add into center rotation or use as an intervention tool"; and to design "mini lessons that focus on the specific type of data that was collected." Such datainformed instruction suggests Monster, P.I. can be a helpful assessment to drive instruction. Overall, this study involved an assessment of the dimensionality of morphology and vocabulary and an assessment of syntax resulting in Monster, P.I., a language assessment that provides scores on three component areas of language (multidimensional morphology and vocabulary and unidimensional syntax) with the goal of informing instruction.

Acknowledgments

We thank the students, teachers, schools, and district that participated in this research. We also thank all research team members who worked on the project, staff Tonya Simmons who kept the project moving, and the Florida Center for Interactive Media team led by Cody Diefenthaler who put together the gaming environment for Monster, P.I.

Declaration of Conflicting Interests

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

Funding

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150199 to Vanderbilt University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

ORCID iD

Sara Jones D https://orcid.org/0000-0002-1423-9974

Supplemental Material

Supplemental material for this article is available on the *Assessment* for *Effective Intervention* website with the online version of this article.

Notes

- 1. This excludes language proficiency assessments that are used at scale with multilingual learners.
- Following Carnegie Council on Advancing Adolescent Literacy (2010), we consider children age 10/Grade 4 and older as adolescents.
- 3. Students participated in different years of the grant with some overlap.
- 4. Morphology was assessed more broadly because the need for the development of a measure of morphology had been emphasized in the literature, whereas assessment of vocabulary was better established (Pearson et al., 2007) and unidimensional syntax was expected. We piloted three additional tasks in Year 1, but data showed the tasks did not add to the model. They were a task where students split words into morphemes, where students chose a nonword with the appropriate suffix to complete a sentence, and where students had to consider morphological meanings in larger passages. We included vocabulary and syntax as guidance from the literature indicated their importance.
- Note that flexMIRT only provides point estimates for indexes in unidimensional models and provides confidence intervals for multidimensional models.

References

- Adlof, S. M., & Hogan, T. P. (2019). If we don't look, we won't see: Measuring language development to inform literacy instruction. *Policy Insights from the Behavioral and Brain Sciences*, 6(2), 210–217.
- Brimo, D., Apel, K., & Fountain, T. (2017). Examining the contributions of syntactic awareness and syntactic knowledge to reading comprehension. *Journal of Research in Reading*, 40(1), 57–74.
- Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, 80(2), 144–179. https://doi.org/10.3102/0034654309359353

- Bowey, J. A., & Patel, R. K. (1988). Metalinguistic ability and early reading achievement. *Applied Psycholinguistics*, 9(4), 367–383.
- Cain, K. (2007). Syntactic awareness and reading ability: Is there any evidence for a special relationship? *Applied Psycholinguistics*, 28(4), 679–694.
- Carlisle, J. F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics*, 9(3), 247–266.
- Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success.* Carnegie Corporation of New York.
- Cho, S., & Goodwin, A. (2017). Modeling learning in doubly multilevel binary longitudinal data using generalized linear mixed models: An application to measuring and explaining word learning. *Psychometrika*, 82(3), 846–870. https://doi. org/10.1007/s11336-016-9496-y
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of educational research*, 86(1), 79–122.
- Crosson, A. C., & Lesaux, N. K. (2013). Does knowledge of connectives play a unique role in the reading comprehension of English learners and English-only students? *Journal of Research in Reading*, 36(3), 241–260.
- Deacon, S. H., & Kieffer, M. (2018). Understanding how syntactic awareness contributes to reading comprehension: Evidence from mediation and longitudinal models. *Journal of Educational Psychology*, 110(1), 72–86.
- Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passagelevel comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, 2(1), 1–44.
- Foorman, B. R., Koon, S., Petscher, Y., Mitchell, A. M., & Truckenmiller, A. J. (2015). Examining general and specific factors in the dimensionality of oral language and reading in grades 4-10: A bi-factor approach. *Journal of Educational Psychology*, 107(3), 884–899. https://doi.org/10.1037/edu 0000026
- Foorman, B. R., Petscher, Y., Stanley, C., & Truckenmiller, A. (2017). Latent profiles of reading and language and their association with standardized reading outcomes in kindergarten through tenth grade. *Journal of Research on Educational Effectiveness*, 10(3), 619–645. https://doi.org/10.1080/19345 747.2016.1237597
- García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, 84(1), 74–111.
- Goodwin, A. P. (2016). Effectiveness of word solving: Integrating morphological problem-solving within comprehension instruction for middle school students. *Reading and Writing*, 29(1), 91–116.
- Goodwin, A. P., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia*, 60(2), 183–208.

- Goodwin, A. P., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, 17(4), 257–285. https://doi.org/10.1080/10888438.2012.689791
- Goodwin, A. P., & Cho, S. J. (2016). Unraveling word learning: Reader and item-level predictors of word learning within comprehension instruction for fifth and sixth graders. *Scientific Studies of Reading*, 20(6), 490–514.
- Goodwin, A. P., Petscher, Y., Carlisle, J. F., & Mitchell, A. M. (2017). Exploring the dimensionality of morphological knowledge for adolescent readers. *Journal of Research in Reading*, 40(1), 91–117.
- Goodwin, A. P., Petscher, Y., Jones, S., McFadden, S., Reynolds, D., & Lantos, T. (2020). The monster in the classroom: Assessing language to inform instruction. *The Reading Teacher*, 73(5), 603–616. https://doi.org/10.1002/trtr.1870
- Goodwin, A. P., Petscher, Y., Reynolds, D., Lantos, T., Gould, S., & Tock, J. (2018). When complexity is your friend: Modeling the complex problem space of vocabulary. *Education Sciences*, 8(4), Article 169.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6–10.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234. https://doi. org/10.3102/0013189X11413260
- Hadley, E., Dickinson, D., Hirsh-Pasek, K., & Golinkoff, R. (2019). Building semantic networks: The impact of a vocabulary intervention on preschoolers' depth of word knowledge. *Reading Research Quarterly*, 54(1), 41–61. https://doi. org/10.1002/rrq.225
- Hart, R., Casserly, M., Uzzell, R., Palacios, M., Corcoran, A., & Spurgeon, L. (2015). Student testing in America's great city schools: An inventory and preliminary analysis. *Council* of the Great City Schools. https://files.eric.ed.gov/fulltext/ ED569198.pdf
- Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing*, 25(2), 347–373.
- Kieffer, M. J., Petscher, Y., Proctor, C. P., & Silverman, R. D. (2016). Is the whole greater than the sum of its parts? Modeling the contributions of language comprehension skills to reading comprehension in the upper elementary grades. *Scientific Studies of Reading*, 20(6), 436–454.
- Ku, Y. M., & Anderson, R. C. (2003). Development of morphological awareness in Chinese and English. *Reading and Writing*, 16(5), 399–422.
- Language and Reading Research Consortium. (2015). The dimensionality of language ability in young children. *Child Development*, 86(6), 1948–1965.
- Logan, J. K., & Kieffer, M. (2017). Evaluating the role of polysemous word knowledge in reading comprehension among bilingual adolescents. *Reading and Writing*, 30(8), 1687– 1704. https://doi.org/10.1007/s11145-017-9745-1
- MacGinitie, W., MacGinitie, R., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie reading test* (4th ed.). Riverside Publishing Company.

- Mitchell, A., & Savill-Smith, C. (2004). The use of computer and video games for learning: A review of the literature. Learning Skills Development Agency.
- Nagy, W., & Anderson, R. (1984). The number of words in printed school English. *Reading Research Quarterly*, 19, 304–330.
- Nagy, W., Carlisle, J. F., & Goodwin, A. P. (2014). Morphological knowledge and literacy acquisition. *Journal of Learning Disabilities*, 47(1), 3–12. https://doi.org/10.1177 /0022219413509967
- Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(1), 91–108.
- Northwest Evaluation Association. (2017). *Measures of Academic Progress Interim Assessments for Grades K–12*. https://www. nwea.org/content/uploads/2015/11/MAP-Teachers-and-School-Leaders-Brochure-FEB17.pdf
- Pacheco, M. B., & Goodwin, A. P. (2013). Putting two and two together: Middle school students' morphological problem solving strategies for unknown words. *Journal of Adolescent* and Adult Literacy, 56(7), 541–553. https://doi.org/10.1002/ JAAL.181
- Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, 42(2), 282–296.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. Scientific Studies of Reading, 11(4), 357–383.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22–37.

- Sireci, S., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement JEM*, 28(3), 237–247. https://doi.org/10.1111/j.1745-3984.1991. tb00356.x
- Spencer, M., Muse, A., Wagner, R., Foorman, B., Petscher, Y., Schatschneider, C., Tighe, E., & Bishop, D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing*, 28(7), 959–988. https://doi.org/10.1007/s11145-015-9557-0
- Taylor, N. A., Greenberg, D., Laures-Gore, J., & Wise, J. C. (2012). Exploring the syntactic skills of struggling adult readers. *Reading and Writing*, 25(6), 1385–1402.
- Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language*, 28, 649–667.
- Tzuriel, D., & George, T. (2009). Improvement of analogical reasoning and academic achievements by the Analogical Reasoning Program (ARP). *Educational and Child Psychology*, 26(3), 71–93.
- Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly*, 50(3), 337–356. https://doi.org/10.1002/rrq.104
- 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.