A peer reviewed, open-access electronic journal: ISSN 1531-7714

The Effects of Stimuli Format on a Computer-Based Decoding Task

Amanda M. Kern, *University of Nebraska at Omaha*
Michelle K. Hosp, *Renaissance*
John Hosp, *University of Massachusetts Amherst*
Amanda M. Marcotte, *University of Massachusetts Amherst*
Derek B. Rodgers, *University of Iowa*

Abstract: The purpose of this study was to examine the effects of test format (i.e., layout of stimuli) on the accuracy and response rate performance of emerging readers on a computer-based nonsense word reading task. Fifty-three second graders completed a modified nonsense word fluency (NWF) task with stimuli presented in five different formats. Format variations included column and row layouts, as well as differences in the number of stimuli. Results from nonparametric comparisons showed there are statistically significant performance differences between test formats for both accuracy and response rate score metrics. Follow-up comparisons show that response rate score metrics appear to be more influenced by test format than accuracy score metrics, and the number of stimuli presented at a time appears to have a significant effect on student performance. Implications for the role of response rate in test development and future research are discussed.

Keywords: Test development, Computer-based testing, Literacy test

Introduction

Reading achievement in the United States remains a persistent concern. Despite the well-documented negative consequences of poor reading proficiency (e.g., increased dropout rates, under- or unemployment), the implementation of national initiatives (e.g., Reading First), and extensive empirical research, proficiency rates remain low. In 2019, prior to the COVID-19 pandemic, only 35% of fourth-grade students and 33% of eighth-grade students scored at or above the proficient level on the National Assessment of Educational Progress (NAEP). These rates were largely unchanged over the preceding decade, indicating a longstanding and unresolved challenge in reading achievement (NCES, 2019). Following pandemic-related disruptions,

proficiency rates exhibited only minimal change, with 33% of fourth graders and 31% of eighth graders scoring at or above proficiency in 2023 (NCES, 2024).

Of greatest concern is that the most substantial declines have occurred among the lowest-performing readers. Students scoring below the 50th percentile experienced 8–10-point decreases in reading performance, with the largest drops among students already performing well below grade level. Combined with the chronic pattern of general low proficiency, these declines underscore the urgent need for instructional practices capable of effectively remediating reading difficulties.

For most students who struggle, the primary skill deficit lies at the word level (Knutson et al., 2004; Shaywitz, 2003) where students struggle to accurately associate phonemes (i.e., speech sounds) with graphemes (i.e., printed letters). As many as 80% of students with a specific learning disability in reading have deficits in word-level skills (Moats & Tolman, 2009). These students often lack efficient word-identification skills, leading to disruptions in fluency and, ultimately, diminished comprehension (Fletcher et al., 2019; Good et al., 2001; Weisner, 2012). Students with such profiles must be identified for targeted intervention.

One method for identifying students in need of intervention is through testing. Increased attention to classroom assessment has coincided with broader adoption of multi-tiered systems of supports (MTSS) and data-based decision-making (DBDM) frameworks, both of which emphasize the use of reliable and valid data to inform instructional planning (Jimerson et al., 2016). Selecting assessments for use within MTSS is therefore a critical decision. In addition to strong psychometric properties, tests must also be efficient and practical. Technology-based assessments are often promoted an avenue to increase efficiency; however, tests that rely solely on reading passages and answering questions do not provide sufficient information about students' decoding skills indicating a need for test development for these foundational skills.

Considerations for Decoding Tests

Without effective intervention, students who struggle with reading in the early grades rarely catch up to their peers (Jimerson et al., 2016; Torgesen, 1998). Reliable, instructionally relevant assessment data are therefore essential for selecting interventions that address specific skill deficits to improve outcomes (Hosp & Ardoin, 2008). Decoding measures, particularly those that analyze grapheme–phoneme correspondence errors, are effective for identifying students at risk for reading difficulties (Li et al., 2025).

Decoding assessments typically require students to read lists of words, though the use of real words versus pseudowords varies across measures (Kern & Hosp, 2018). Pseudowords (e.g., *mip*, *zilk*) minimize the influence of sight-word memory and provide a more direct measure of decoding because they cannot be recognized as whole units. Research continues to demonstrate that pseudoword tasks isolate phonological decoding more effectively than real word reading, which often draws on stored lexical knowledge (Zhang & Peng, 2022). Pseudoword reading also reflects students' familiarity with orthographic patterns and predicts real word reading and overall reading proficiency (Carver, 2003; Pierce et al., 2010). More recent findings reinforce the instructional relevance of pseudowords with text interventions producing measurable gains in pseudoword decoding (Odo, 2024).

Decoding assessments should also capture response rate (i.e., the speed and automaticity with which students apply decoding skills). Students who decode words quickly and accurately in isolation are more likely to transfer these skills to connected text (Ehri, 1980). Recent fluency research supports reading fluency as an independent construct distinct from reading accuracy strongly predicts reading comprehension (Hsu et al., 2023). Accordingly, assessments that measure only accuracy may underestimate the proficiency required for skilled reading.

Traditional timed decoding assessments limit the number of items students can attempt, reducing opportunities to observe the full range of their decoding skills (Kern & Hosp, 2018). This constraint is especially problematic for struggling readers, who require more time per item (Ritchey, 2008; Reutzel et al., 2014). Although accuracy is important, slow and effortful decoding, even when correct, does not reflect the automaticity needed for proficient reading (Joshi & Aaron, 2002). Emerging assessment frameworks increasingly integrate both accuracy and rate to generate more comprehensive skill metrics and increase accurate risk identifications (Huang et al., 2025).

An alternative to timed assessments is to record the total time needed for a student to read an entire word list and calculate response rate by dividing words read by time elapsed. This method avoids limiting the number of items but may be time-consuming. Decoding assessments should therefore balance the need for sufficient skill demonstrations with practical testing constraints. Optimized test formats that efficiently measure both accuracy and rate offer the most practical solution.

Test Format Considerations in Decoding Assessment

The visual layout of decoding assessments is an important but often overlooked factor in maximizing testing efficiency. Historically, decoding tests have changed little in format. Most require students to identify words in isolation, typically arranged in rows or columns. Curriculum-based measures (CBMs) of nonsense word fluency (NWF), such as DIBELS, present pseudowords in horizontal rows on a single page, whereas standardized assessments including the *Word Attack* subtest of the *Woodcock Reading Mastery Tests* (WRMT) and the *Phonemic Decoding Efficiency* subtest of the *Test of Word Reading Efficiency* (TOWRE) typically organize items in vertical columns. Despite these differences, little empirical work has examined how such formats influence the measurement of decoding accuracy or response rate. As a result, test formats generally reflect developer preference and are replicated across test editions.

Word lists are the most common presentation format for decoding assessment, and empirical work supports their diagnostic utility. Jenkins et al. (2003) found that list reading is a sensitive method for identifying reading impairments. Skilled readers consistently outperformed impaired readers in both contextual and list formats, and all students who demonstrated impairment during contextual reading also showed impairment during list reading.

However, test format and item arrangement may interact with visual-perceptual factors that influence reading. Jones et al. (2009) demonstrated that item layout in a letter-identification task affected performance differently for skilled and less skilled readers: proficient readers benefited from larger item arrays, likely due to rapid processing and effective preview of upcoming items, whereas struggling readers did not. Similar patterns appear in eye-movement research. Chace et al. (2005) showed that skilled readers exhibit parafoveal preview benefit (i.e., faster recognition of a word due to pre-processing of the upcoming word) while unskilled readers do not, consistent with automaticity theories that unskilled readers must devote full cognitive resources to the fixated word (Perfetti, 1986).

More recent studies highlight how visual constraints may influence decoding performance across different item layouts. Chiu and Drieghe (2023) demonstrated that visual crowding substantially reduces the efficiency of parafoveal processing, suggesting that tightly spaced or visually dense test formats could hinder some readers' performance. Similarly, Frömer et al. (2015) found that individual differences in extrafoveal acuity and susceptibility to crowding were associated with reading speed and the magnitude of the preview benefit. Readers with stronger extrafoveal acuity and lower crowding sensitivity were more effective at using parafoveal information.

In addition, recent work has shown that the preview benefit is not restricted to horizontal reading. Al Jassmi et al. (2025) examined parafoveal processing in vertical alphabetic reading and found clear preview-benefit effects in this format, though the magnitude and distribution of the benefit differed from horizontal

reading. Their findings suggest that while readers can extract parafoveal information in vertically presented text, the efficiency and spatial characteristics of preview processing may shift depending on orientation. This has direct implications for decoding assessment: vertical item arrangements (such as those used in the WRMT or TOWRE) may alter how easily readers can access upcoming items, particularly given individual differences in extrafoveal acuity and crowding sensitivity documented by Frömer et al. (2015).

Taken together, these findings indicate that decoding test formats may advantage or disadvantage readers depending on their visual-perceptual skills and the extent to which they benefit from parafoveal preview. Skilled readers, who show stronger preview benefit and lower susceptibility to crowding, may perform differently on horizontal versus vertical item layouts, whereas struggling readers, who typically show reduced preview benefit, may be less affected by orientation and spacing. These visual-processing considerations highlight the need for investigation into how decoding test formats influence performance.

Technology Test Platforms and Format Considerations

An additional dimension of test format research concerns the role of technology platforms in assessment administration. Access to digital devices in schools has increased substantially, with 96% of public schools reporting providing devices to their students (NCES, 2023). Computer-based tests (CBTs) are now widely used, particularly in large-scale assessment programs such as the National Assessment of Educational Progress (NAEP). More recently, technology has also been integrated into formative and classroom-based assessments, where it is viewed as a means of enhancing efficiency and instructional utility (Brown et al., 2008). Broad reviews of CBT adoption note that while digital platforms offer opportunities for streamlined administration, automated scoring, and adaptive item delivery, they also introduce considerations related to digital literacy, mode comparability, and student experience (Panzarella & Walmsley, 2025; Walker & Handley, 2022).

Several curriculum-based measurement (CBM) publishers, including DIBELS, FastBridge Learning, and Aimsweb+, now offer technology-enhanced administration options for reading assessments (Hosp et al., 2016). However, these digital versions typically replicate the traditional paper-based format, presenting identical items arranged in rows or columns on a screen. Research on test adaptation cautions that simply transferring paper tests into digital environments without redesign may fail to leverage technological affordances and may also introduce construct-irrelevant variance or mode effects if the digital format does not support comparable cognitive processing demands (Lynch, 2022). Cognitive load studies in digital versus paper testing settings suggest that CBT administration can alter the mental demands associated with navigating and responding to assessment items, with implications for how students engage with test items (Whipp & Malpique, 2024).

Given evidence that CBTs can improve testing efficiency, it is essential to examine not only the mode of delivery but also the format of the stimuli presented within these platforms. Understanding how layout, spacing, orientation, and the number of items displayed influence performance may reveal opportunities to enhance measurement precision while maintaining reliability and validity standards. Technology platforms offer opportunities for responsive design, adjustable typography, or alternative display formats that could better align with students' visual processing strengths, features that traditional paper tests cannot accommodate. In their 2022 analyses of digital assessments, Walker & Handley emphasize the importance of such design considerations, noting that well-designed CBTs can improve engagement and accessibility, but poorly designed ones may compromise score interpretations.

Student format preferences, such the number of items per screen or whether words are arranged in rows versus columns, may further inform test design decisions. Integrating such considerations with research on list reading and visual processing may lead to assessment formats that optimize student performance and reduce the time required for administration. Ultimately, because teachers rely on assessment data to make

timely instructional decisions, test formats that improve efficiency and engagement without compromising validity may enhance instruction.

Purpose of This Study

The purpose of this present study was to determine if test format affected decoding accuracy and response rate of second grade students on a computer-based nonsense word reading task. Comparisons of interest included differences between column and row layouts of stimuli, differences in the number of stimuli presented at one time, and differences between participants indicated preferred format and other formats. We hypothesized that there would be performance differences across formats and based on the results of eye tracking studies and the presence of preview benefit, that row formats would optimize reading performance. We also hypothesized that students would perform optimally on their preferred format, indicating a relation between preference and performance.

Methods

Participants

Participants in this study were students in Grade 2 (n=53), recruited from three second-grade classrooms from one school, in a small city in a northeastern state in the United States. In the sample, 43% were male, 83% were white, 3.8% were African American, 1.8% were Asian, 5.7% were Latino/Hispanic, and 5.7% were multi-race, non-Hispanic. In addition, 5.7% of participants were English Language Learners, 26.4% were receiving Title One Reading Services, and 1.8% were students with disabilities. Parental consent and participant assent were collected prior to data collection, and no participation incentive was provided.

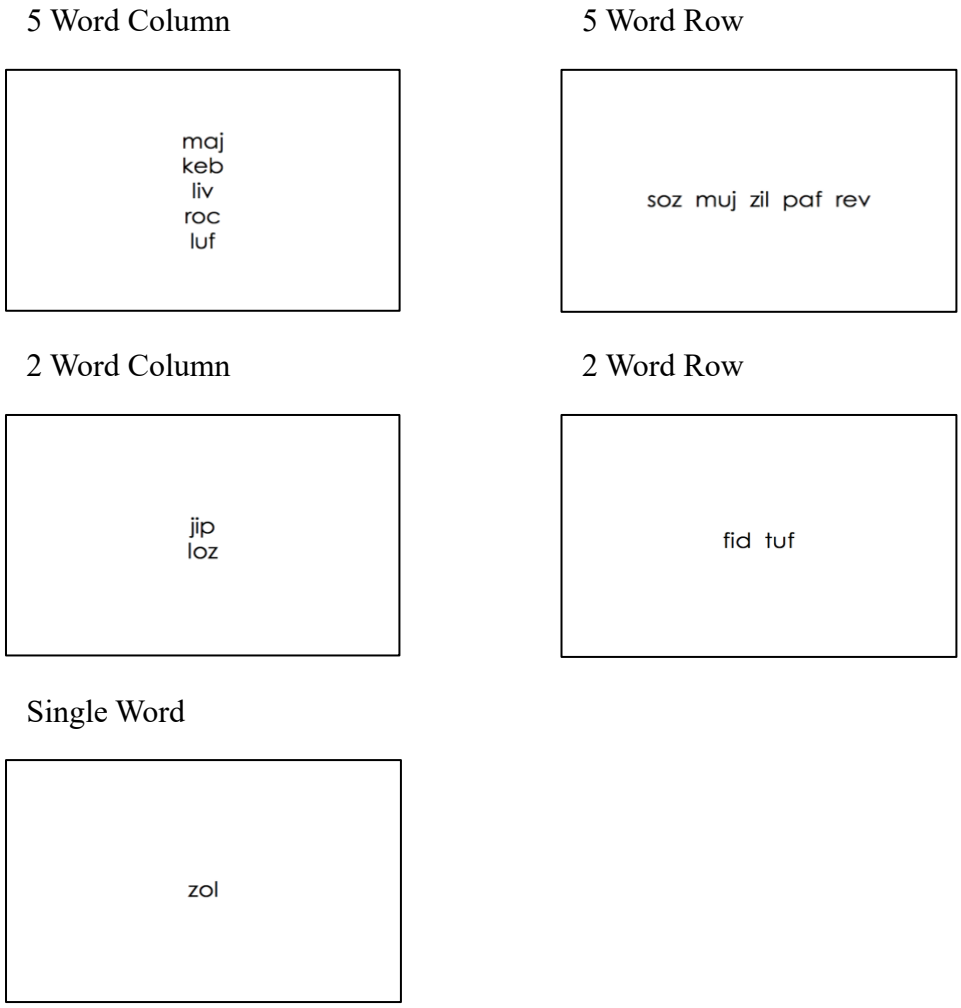
Instrument

Participants in this study completed a computer-based modified NWF, created for this study by the primary researcher. The task was developed after reviewing common decoding tests and reviewing decoding learning standards of the participants (i.e., second grade learning standards). The task included a total of 100 unique consonant-vowel-consonant (CVC) nonsense words divided equally between 5 format conditions (i.e., 20 words per format). The five format conditions were: 5-word column (5C), 5-word row (5R), 2-word column (2C), 2-word row (2R) and a single word (S) presentation. Examples of the five formats can be found in Figure 1. To help ensure format equivalency each simple short vowel (i.e., a,e,i,o, and u) was represented an equal number of times in each condition. In addition, within each format, and within each vowel category, no beginning or ending consonants were repeated (e.g., within the 5C format there were four CVC nonsense words with "a" as the vowel, and across those four words there are no repetitions of beginning or ending consonants).

Directions were provided to the student at the onset of the task. The data collectors, which included the primary researcher as well as trained graduate students, controlled the advancement of the slides using the computer mouse. Scoring was done by giving credit for each correct letter-sound correspondence (CLS) that was identified. In addition, scores were also calculated for each whole word read (WWR). Students received credit for each WWR by correctly blending all correct phonemes. This scoring procedure matches standard scoring instructions for most common NWF CBMs. Both CLS and WWR scores, and completion time were recorded for each format. Time was used to determine a student's response rate (CLS/minute and WWR/minute) on each test format conditions. All scores were used in analyses to determine differences in student performance across test formats. To confirm scoring reliability, data collection sessions were recorded and 20% of protocols were rescored. Scoring agreement was 98%.

The task was presented to the student using Microsoft PowerPoint. The 5-condition test format task included a total of 48 slides. Each slide had a white background with black text. Text was centered on each slide, using Century Gothic, 42-point font. Columns were single spaced, and there was a double space between each word in row formats. To control for practice effects, the presentation order of conditions was counterbalanced, and each participant was given a different presentation order of formats.

Figure 1. Examples of the Five Test Formats



Preference Indication

Following the completion of the five-condition test format task, each student was asked to indicate which test format they preferred most. The final slide of the PowerPoint presentation presented thumbnail images of each of the format conditions. Students were asked to indicate which of the formats they liked best. Students verbally indicated or pointed to which format they preferred.

Analyses

Based on the sample size of this study (n=53) and the number of comparisons of interest, non-parametric analyses were performed. The Friedman test (Friedman, 1937) was run to determine if there were performance differences between test formats for each of the score metrics. Effect sizes are reported using

Kendall's W, and follow-up comparisons using Wilcoxon signed-rank tests. All analyses were completed using R Statistical Computing software (R Core Team, 2019).

Results

Descriptive statistics for all formats and score metrics are presented in Table 1. The highest mean score for each dependent variable in the test format condition (i.e., CLS, WWR, CLS/min, and WWR/min) occurred in the 5C condition. The lowest mean scores for accuracy (i.e., CLS and WWR) occurred in the 5R condition. The lowest mean scores for response rate (i.e., CLS/min and WWR/min) occurred in the S condition.

Table 1. Descriptive Statistics

Score Metric & Format	M	SD	Minimum	Maximum	Skewness
CLS					
5C	53.19	5.63	38.0	60.0	-0.87
5R	51.49	6.03	37.0	59.0	-0.73
2C	52.15	5.85	30.0	60.0	-1.56
2R	52.72	5.61	39.0	60.0	-0.98
S	52.85	5.22	38.0	59.0	-1.13
WWR					
5C	12.91	6.20	0	20.0	-0.86
5R	11.34	5.78	0	19.0	-0.54
2C	12.04	5.75	0	20.0	-0.82
2R	12.13	6.29	0	20.0	-0.93
S	12.30	6.08	0	19.0	-0.95
CLS/min					
5C	67.16	34.40	9.6	165.7	0.73
5R	63.63	29.31	20.6	139.2	0.73
2C	53.82	23.58	8.9	105.0	0.28
2R	54.23	22.17	11.7	99.4	0.07
S	46.17	17.49	10.9	82.5	-0.03
WWR/min					
5C	17.95	13.27	0	51.4	0.55
5R	15.32	11.18	0	43.2	0.66
2C	13.39	9.13	0	31.9	0.24
2R	13.63	9.03	0	30.9	0.06
S	12.13	8.71	0	47	1.03

Note. CLS= correct letter sounds; WWR= whole words read

Differences across Format Conditions

Correct letter sounds. A Friedman test showed that for this sample the differences in CLS scores between the five test format conditions were not statistically significant, $\chi 2 = 8.53$, df= 4, $p = .07$, Kendall's W =0.04. Given no statistically significant differences were noted for CLS scores across the test format conditions no follow-up analyses were performed.

Whole words read. A Friedman test showed that for this sample the differences in WWR performance between the five test format conditions were statistically significant, $\chi 2 = 14.39$, df=4, $p = .006$, Kendall's W =0.07 (small effect). To determine the differences in WWR between each test format conditions follow-up

pairwise comparisons using the Wilcoxon signed-rank test (with Bonferroni adjustments to account for the number of comparisons) were completed. Results show that one comparison was significant. WWR scores in the 5C condition were significantly greater than scores in the S condition (W=926, adjusted p = .005).

Correct letter sounds per minute. A Friedman test showed that for this sample the differences in CLS/min performance between the five test format conditions were statistically significant, $\chi 2$ = 73.75, df = 4, p<.0005, Kendall's W =0.35. To determine the differences in CLS/min between each test format conditions follow-up pairwise comparisons using the Wilcoxon signed-rank test (with Bonferroni adjustments to account for the number of comparisons) were completed. Results show eight significant comparisons. Scores in the 5C condition were significantly greater than scores in the 2C, 2R, and S condition (W = 148-1338, adjusted p values all < .0005). Scores in the 5R condition were significantly greater than scores in the 2C, 2R, S conditions (W = 258-1347, adjusted p values all ≤ .004). Scores in the 2C condition were significantly greater than in the S condition (W= 1136, adjusted p =.002). Last, scores in the 2R condition were significantly greater than in the S condition (W= 1234, adjusted p< .005).

Words read correctly per minute. A Friedman test showed that for this sample the differences in WWR/min performance between the five test format conditions were statistically significant, $\chi 2$= 33.89, df= 4, p<.0005, Kendall's W = 0.16. To determine the differences in WWR/min between each test format conditions follow-up pairwise comparisons using the Wilcoxon signed-rank test (with Bonferroni adjustments to account for the number of comparisons) were completed. Results show four significant comparisons. Scores in the 5C condition were significantly greater than scores in the 2C, 2R, and S conditions (W range = 117-1114, adjusted p values all <.0005). Score in the 5R condition were significantly greater than scores in the S condition (W=1050, adjusted p = .01).

Differences Based on Format Preference

Participants were asked to indicate their preference for format. Nineteen participants selected the S format, fifteen selected 5C, nine selected 5R, six selected 2C, and four selected 2R. In the sample, 24.5% (n=13) of participants selected the format on which their CLS and CLS/min scores were highest, and 22.6% (n=12) of participants selected the format on which their WWR and WWR/min scores were highest.

The differences in all score metrics between the indicated preferred format and nonpreferred formats were all examined via the Friedman test. Across all score metrics, there were no statistically significant differences between scores on the indicated preferred format and all others (CLS: $\chi 2$= .98, df= 4, p = .91; WWR: $\chi 2$= 3.02, df= 4, p = .55; CLS/min: $\chi 2$= 4.43, df= 4, p = .35; WWR/min: $\chi 2$= 1.93, df= 4, p = .75) These results suggests that student preference for format had no significant effect on any of the score metrics.

Discussion

Effective assessment practices positively influence student outcomes (Black & Wiliam, 2009), and teachers who use valid assessment data make more accurate instructional decisions (Hosp & Ardoin, 2008). Because decoding assessments provide essential diagnostic information about students' word-reading skills, it is important to evaluate whether test format features influence performance.

List-reading tasks are sensitive to reading impairments (Jenkins et al., 2003). The present study examined whether *format* variations in such tasks affect decoding accuracy and response rate. Participants completed a modified NWF task across five formats (5C, 5R, 2C, 2R, S), with accuracy (CLS, WWR) and response-rate metrics (CLS/min, WWR/min) computed for each. Performance varied by score type and format, with significant differences observed for three of the four metrics. Effects were more pronounced for response-rate measures (Kendall's W = .35 for CLS/min; .16 for WWR/min), suggesting that format influences

the efficiency with which students apply decoding skills more than their accuracy. Because fluency reflects automatic and efficient decoding, response-rate metrics may provide particularly valuable information about skill mastery.

The number of words displayed at one time influenced performance: students showed significantly higher response rates on the 5C and 5R formats than on the 2C, 2R, and S formats. These findings align with eye-movement and parafoveal processing research showing that readers process upcoming words while fixating the current one, resulting in faster reading when multiple words are visible (Chace et al., 2005; Al Jassmi et al., 2025). In contrast, single-word presentations eliminate opportunities for parafoveal preview.

In this sample, no significant differences emerged between column and row layouts (2C vs. 2R; 5C vs. 5R), indicating that orientation did not meaningfully affect performance. Notably, the highest scores occurred in the 5C format rather than the more traditional 5R layout. This may reflect differences associated with computer-based versus paper-based presentation modes; research on mode effects shows that simply transferring test content from paper to computer can introduce construct-irrelevant variance and alter performance (Lynch, 2022). In addition, digital formats may impose different cognitive demands during reading and item navigation, influencing how students process assessment stimuli (Whipp & Malpique, 2024). Further research is needed to determine how these mode-related factors contribute to decoding performance and to guide optimal design for digital decoding assessments.

Although previous research suggests that assessment preference can influence performance (e.g., Jones et al., 2015), the present study found no significant differences between students' preferred formats and their performance on non-preferred formats (χ^2 range = .98–4.43). Performance was relatively consistent across formats. We consider several possible reasons for this result. Students in this sample may have possessed sufficiently strong decoding skills to perform similarly across conditions, limiting variation. Younger students may also lack metacognitive awareness of how format affects their performance. Informal comments suggested that some students perceived multi-word formats as faster, whereas others preferred single-word displays, indicating subjective but not performance-related differences.

These results differ from studies showing benefits of student choice for motivation and performance (e.g., Randi & Corno, 2000; Ciampa, 2016). However, participants in the present study did not choose the format in which they were assessed; they selected a preferred format only after completing all conditions. Allowing students to select their preferred format before the assessment may produce different outcomes. Future research should investigate whether incorporating choice improves performance or efficiency in decoding assessments.

Implications for Test Development

Although the content of this study was specific to decoding, the results may have implications for broader educational test development. Test format, and the presentation of stimuli, are likely variables that should be considered during development. As more traditional paper tests convert to CBTs it is important to consider how items are presented to test-takers. This study indicated that student's accuracy and response rates were impacted based on number of items displayed at a time as well as item arrangement (i.e., vertical versus horizontal). It should not be assumed that moving a previous paper version onto a computer screen will result in the same test scores. Mode effects should be analyzed, the organization of stimuli, and the number of items presented are all factors to be considered. Given previous research on preview benefit in optimizing reading, the researchers hypothesized that row presentations would be optimal.

A final implication for consideration regards student preference for test format. We hypothesized that performance would be optimized on the preferred format of the participants but found no significant differences between performance on selected preference compared to other formats. Test developers may apply this result in two ways. First, given that no significant differences were found, a single format may be

developed and used for a test (i.e., no variety in formats are available). Second, test developers may consider developing multiple formats for tests. Each format must meet standards for reliability and validity, but it may be that for students resistant to the testing task, that the options for choice in formatting may encourage participation, making the testing experience less aversive. These considerations warrant attention in future research studies.

Limitations

There are several limitations to this study that should be noted. First, this is a preliminary study, and therefore the small sample size and demographic limitations make it difficult to generalize results. We used nonparametric tests in our analyses to build confidence in our results with the limitations of our sample size. While our preliminary findings suggest test format may have effects on performance, future studies with a larger sample are needed to confirm. Second, this study did not examine or account for the decoding instruction that participants were receiving. Novelty of task (i.e., reading nonsense words) and level of familiarity with decoding may have influenced the performance measured. Future research should address questions on potential confounds that may impact performance differences such as reading abilities (i.e., format differences between high and low achieving students), and other individual differences, and examine effects of instructional content.

Received: 2/25/2025. **Accepted:** 1/2/2026. **Published:** 1/15/2026.

Citation: Kern, A. M., Hosp, M. K., Hosp, J., Marcotte, A. M., & Rodgers, D. B. (2026). The effects of stimuli format on a computer-based decoding task. *Practical Assessment, Research, & Evaluation*, 31(1)(4). Available online: https://doi.org/10.7275/pare.2924

Corresponding Author: Amanda M. Kern, University of Nebraska at Omaha, Omaha.
Email: akern@unomaha.edu

References

Al Jassmi, M. A., Warrington, K. L., McGowan, V. A., Xie, F., & Paterson, K. B. (2025). Parafoveal preview benefit effects in vertical alphabetic reading. *Language, Cognition and Neuroscience*, 1–10. https://doi.org/10.1080/23273798.2025.2327379

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Carver, R. P. (2003). The highly lawful relationships among pseudoword decoding, word identification, spelling, and reading comprehension. *Scientific Studies of Reading, 7*(2), 127–154. https://doi.org/10.1207/S1532799XSSR0702_3

Chace, K. H., Rayner, K., & Well, A. (2005). Eye movements and phonological parafoveal preview: Effects of reading proficiency. *Memory & Cognition, 33*(1), 121–129. https://doi.org/10.3758/BF03195304

Chiu, T. Y., & Drieghe, D. (2023). The role of visual crowding in eye movements during reading: Effects of text spacing. *Attention, Perception, & Psychophysics, 85*(8), 2834–2858. https://doi.org/10.3758/s13414-023-02796-4

Ciampa, K. (2016). Motivating grade 1 children to read: Exploring the role of choice, curiosity, and interest. *Journal of Research in Childhood Education, 30*(1), 14–29.

Ehri, L. C. (1980). The development of orthographic images. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 311–338). Academic Press.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2019). *Learning disabilities: From identification to intervention* (2nd ed.). Guilford Press.

Frömer, R., Dimigen, O., Niefind, F., Krause, N., Kliegl, R., & Sommer, W. (2015). Are individual differences in reading speed related to extrafoveal visual acuity and crowding? *PLOS ONE, 10*(3), e0121986. https://doi.org/10.1371/journal.pone.0121986

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*(3), 257–288. https://doi.org/10.1207/S1532799XSSR0503_4

Hosp, M. K., & Ardoin, S. P. (2008). Assessment for instructional planning. *Assessment for Effective Intervention, 33*(2), 69–77. https://doi.org/10.1177/1534508407311428

Hsu, C. C., Chan, D. W., & Ho, C. S.-H. (2023). Examining the relationship between reading fluency and reading comprehension in Chinese. *Frontiers in Psychology, 14*, 1221396. https://doi.org/10.3389/fpsyg.2023.1221396

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2016). *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.). Springer.

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*(4), 719–729. https://doi.org/10.1037/0022-0663.95.4.719

Jones, A. C., Antonacci, P., & Perkins, K. (2009). Effects of matrix size and item arrangement on letter identification performance. *Journal of Behavioral Education, 18*(2), 97–115. https://doi.org/10.1007/s10864-008-9070-9

Joshi, R. M., & Aaron, P. G. (2002). Rate and accuracy in reading: Are they the same or different? *Reading and Writing, 15*, 593–611. https://doi.org/10.1023/A:1020968205648

Kern, A. M., & Hosp, M. K. (2018). The status of decoding tests in instructional decision-making. *Assessment for Effective Intervention*, *44*(1), 32-44. https://doi.org/10.1177/1534508417749874

Li, Y., Kim, Y., & Ahmed, S. (2025). The utility of decoding measures for identifying at-risk readers: A meta-analysis. *Educational Psychology Review, 37*(1), 55–80.

Moats, L. C., & Tolman, C. (2009). *Language essentials for teachers of reading and spelling*. Sopris West.

National Center for Education Statistics. (2019). *2019 NAEP reading assessment highlights: Reading results for grades 4 and 8* [NAEP Report Card]. U.S. Department of Education. https://www.nationsreportcard.gov/highlights/reading/2019/

National Center for Education Statistics. (2023). *Elementary and secondary education: Technology support for students and families*. https://nces.ed.gov/surveys/annualreports/topical-studies/covid/theme/elementary-and-secondary-education-technology-support/

National Center for Education Statistics. (2024). *NAEP reading assessment results: National achievement-level results* [Data set]. National Assessment of Educational Progress (NAEP). U.S. Department of Education. https://www.nationsreportcard.gov/reading/nation/achievement/

Odo, D. (2024). Decodable texts improve early reading outcomes: A meta-analysis. *Journal of Literacy Research, 56*(2), 145–168.

Perfetti, C. (1986). Continuities in reading acquisition, reading skill, and reading disability. *Remedial and Special Education, 7*(1), 11–21. https://doi.org/10.1177/074193258600700104

Pierce, M. E., Katzir, T., Wolf, M., & Fried, M. (2010). The effect of orthographic patterns on word recognition. *Annals of Dyslexia, 60*(1), 1–20. https://doi.org/10.1007/s11881-009-0028-1

Reutzel, D. R., Petscher, Y., & Anderson, D. (2014). Exploring the value of timed and untimed oral reading fluency measures. *Reading and Writing Quarterly, 30*(1), 93–119. https://doi.org/10.1080/10573569.2013.857971

Ritchey, K. D. (2008). The relationship between decoding and fluency: Developmental evidence. *Reading Research Quarterly, 43*(3), 342–362. https://doi.org/10.1598/RRQ.43.3.4

Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. Knopf.

Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator, 22*(1–2), 32–39.

Zhang, H., & Peng, P. (2022). What's in a word? A meta-analysis of pseudoword and real-word reading in developmental dyslexia. *Annals of Dyslexia, 72*(2), 349–370. https://doi.org/10.1007/s11881-022-00259-y