

1. SIGNIFICANCE

Overall Importance of the Problem

The educational and societal cost of illiteracy is enormous. Adults with low literacy are less likely than adults with average or above average literacy achievement to graduate high school, attend college, and be employed full-time (Kutner et al., 2007), as well as more likely to live in poverty, be incarcerated and report poorer overall health (Goodman et al., 2013). Low literacy has a substantial impact on national economic welfare; if literacy skills of the US labor force increased by merely $\frac{1}{4}$ of a standard deviation, the GDP would increase by over 40 trillion dollars by 2090 (OECD, 2010).

Literacy failure in the US is epidemic. The most recent National Assessment of Educational Progress (NAEP) reports revealed that only 35% of fourth graders and 34% of eighth graders read at a proficient or higher level, and students' reading performance was lower in 2019 than 2017, suggesting that currently available solutions are not working (US Department of Education, 2019). Writing outcomes are even worse than reading outcomes. Only 27% of eighth and twelfth graders write at a proficient or higher level (US Department of Education, 2012). Likewise, teachers report that almost 30% of their students experience difficulty with spelling, despite their use of research-supported practices (Graham et al., 2008). Research that leads to effective remediation of these literacy difficulties is vital.

Inadequate knowledge of the English spelling system interferes with higher-level literacy tasks. According to the Not-So-Simple View of Reading (Joshi & Aaron, 2000), one's ability to derive meaning from text depends on the ability to *recognize words* and to *comprehend the language of the text*; and that both of these processes must happen *fluently* for optimal meaning derivation. Inadequate knowledge of the English spelling system has a direct negative impact on word recognition and reading fluency, and an indirect negative impact on reading comprehension (Clemens, Oslund, Simmons, & Simmons, 2014; Conrad, Harris, & Williams, 2013; Ehri & Wilce, 1987; Ehri, 1989; Price, Meisinger, Louwerse, & D'Mello, 2016; Schwanenflugel et al., 2006; Uhry & Shepherd, 1993). Therefore, spelling instruction is vital for students' literacy, and subsequent occupational, success.

Inadequacy of Current Typical Practice

Many teachers do not have strong knowledge of the English spelling system and report low confidence in their ability to teach English spelling effectively (e.g., Adoniou, 2014; Bos, Mather, Dickson, Podhajski, & Chard, 2001; Cunningham, Zibulsky, & Callahan, 2009; McCutchen, Harry, Cunningham, Cox, Sidman, & Covill, 2002; Moats, 1994; Piasta, Connor, Fishman, & Morrison, 2009; Spencer, Schuele, Guillot, & Lee, 2008). As a result, current typical classroom practice in the instruction of the English spelling system is inadequate to meet the literacy needs of all students. Preliminary data from our lab suggests that the majority of teachers (80%) believe that spelling is not adequately addressed in the classroom (Werfel, Dubinsky, & Buell, unpublished data). We propose three reasons for this:

Issue 1: Although spelling is a linguistic skill, typical instruction of English spelling is memory- rather than language-based. The most common instructional approach teachers report utilizing is memorizing words (Fresch, 2007). In the memorization-based view, words are added one at a time to the spelling lexicon (e.g., the weekly spelling list). However, research over the past several decades has confirmed that words are added systematically to the spelling lexicon via an understanding of their linguistic features (Apel, Wilson-Fowler, Brimo, & Perrin, 2012; Deacon, Kirby, & Casselman-Bell, 2009; Werfel, Schuele, & Reed, 2019; Wolter & Apel, 2010). In a recent meta-analysis, Weiser and Mathes (2011) reported robust positive effects of linguistic-based instruction on spelling outcomes. Despite the strong theoretical and empirical linguistic basis for spelling, however, little to no time is spent developing and applying linguistic skills in spelling instruction (Carreker, Joshi, & Boulware-Gooden, 2010; Coltheart & Prior, 2007; Schlagal, 2002).

Issue 2: Teachers have insufficient training in the linguistic principles of English spelling. Teachers have difficulty with linguistic concepts such as counting the number of sound segments in words (e.g., Spencer et al., 2008), counting the number of morphemes in words (e.g., Moats, 1994), and understanding the rules that map sounds to letter and letters to sounds (e.g., Bos et al., 2001). These difficulties are related to lower spelling skills of their students (Puliatte & Ehri, 2018). Teachers report frustration with not only their students' low spelling abilities, but also with their own limited ability to provide effective spelling instruction (Johnston, 2001). Preliminary data indicates that 73% of teachers believe their training programs did not prepare them to provide high-quality spelling instruction (Werfel et al., unpublished data).

Issue 3: Teachers report limited use of adaptations for struggling and/or gifted learners. Adaptations in spelling instruction tailored to that student's unique needs is a primary characteristic of effective spelling instruction (Wallace,

2006). Despite this, almost half of teachers report making virtually no adaptations in their implementation of spelling instruction to accommodate the skill levels of individual students (Graham et al., 2008). There is a great unmet need, therefore, for an efficient method of teaching children the linguistic principles of the English spelling system and individualizing classroom spelling instruction to student skill levels, one for which current resources and teaching methods do not allow.

Computerized Spelling Instruction: A Solution

Computerized spelling instruction provides a solution for these three issues. In computerized spelling instruction, relevant linguistic principles can be embedded within the programming. The result is the provision of accurate, evidence-based instruction, even when teachers do not have explicit knowledge of the linguistic system. Additionally, computerized spelling instruction can be programmed to adapt to an individual student's spelling skill level. The proposed project uses one such computerized spelling instructional program, *Wordification*, to determine the optimal doses of spelling instruction for three levels of spelling ability: strong (above average), average, and poor (below average). *Wordification*, described in detail in the **Approach** section, was developed by the PIs based on Gee's (2003) learning principles for educational games, as well as data-based principles of spelling instruction (Schlagal, 2002; Wallace, 2006) and theoretical accounts of spelling acquisition (Apel, Masterson, & Hart, 2004; Bourassa & Treiman, 2001; Ehri, 1986; Gentry, 1978; Henderson, 1990).

Need for a Phase 1 Trial to Determine Adequate Dose of Spelling Instruction

It is unclear from the previous literature how many exposures to instruction of a spelling pattern are needed to master the skill. Although almost 100% of teachers report teaching spelling, the time spent on spelling instruction in typical elementary school classrooms varies tremendously (Graham et al., 2008); teachers devote, on average, 90 minutes per week to spelling instruction, or over 20 minutes per day in a traditional 4 days instruction + Friday spelling test model. The standard deviation, however, was over 70 minutes, indicating remarkable variation across teachers. Some teachers reported teaching spelling for as little as two minutes per week, while others exceeded three hours per week.

Further, it is likely that the number of exposures to instruction of a spelling pattern needed for mastery differs across student spelling skill groups (strong, average, poor). First, when teachers do not use modifications for different groups of spellers, strong spellers make adequate gains across a school year, whereas poor spellers make few gains (Morris, Blanton, Blanton, & Perney, 1995). Additionally, gains for poor spellers have been reported when supplemental intervention is provided in addition to classroom instruction (Graham, Harris, & Chorzempa, 2002). Finally, in other linguistic-based skills, such as word-learning, it is well-established that a larger dose is needed for those students who have poorer skills (e.g., Storkel et al., 2017). Taken together, it appears that poor spellers require a greater number of exposures to instruction of a spelling pattern to make gains, but the adequate dose of spelling instruction for each skill level of speller is unclear.

Preliminary Data

In a preliminary study of *Wordification*'s effectiveness in increasing spelling ability in elementary school students, we found that, overall, students' spelling of targeted patterns improved over a two- to four-week intervention that occurred for approximately 15 minutes, 4 days per week. Tau-U of 0.66 ($p < .001$) indicated a large change. When we examined outcomes for strong, average, and poor spellers separately, however, clear patterns of improvement occurred. Strong spellers had a Tau-U of 1.0 ($p < .01$), which indicated a very large change, average spellers had a Tau-U of 0.81 ($p < .01$), which indicated a very large change, and poor spellers had a Tau-U of 0.26 ($p = 0.41$), which indicated no or a small change. Therefore, we propose in the current application to separately determine adequate treatment intensity for these three groups of students.

2. INNOVATION

As described above, current educational practice in spelling instruction does not align with evidence for best practice. The proposed research is innovative because it utilizes a behind-the-scenes database to teach linguistic principles of English spelling that are not well-understood by many teachers. The **aims** are innovative in that they address a problem that represents a major gap in extant literature. Although there is much evidence that literacy failure in the US is epidemic and that linguistic-based spelling instruction is optimal for improving outcomes, research that has compared dose and treatment moderators for spelling instruction is rare. Results of this line of inquiry will provide immediate applicability to educational practice. The **approach** is also innovative. *Wordification* is a web-based game system driven by an underlying database that encodes phonological and graphemic features of individual words together with each word's spelled form and examples of the word in sentential context. One fundamental innovation of *Wordification* is that the teaching of the linguistic principles of the English spelling system is embedded within the programming. The application incorporates, hidden to the user,

theoretical approaches to spelling acquisition and data-based spelling instructional strategies. Therefore, the delivery of classroom spelling instruction through *Wordification* will provide accurate instruction of the linguistic principles of English spelling, even when teachers do not have explicit understanding of them. Another key innovative component of *Wordification* is the architecture of the database. Each word's entry contains its orthographic representation, its phonemes and graphemes (mapped in one-to-one correspondence), and example sentences illustrating its meaning in context. Each word's database record is coded so as to point to and play audio files containing the word's full pronunciation and its constituent phonological components.

3. APPROACH

The proposed project will be a Phase 1 dose-finding study, utilizing a slope escalation design based on nontoxic chemotherapeutic agent trials (Hunsberger, Rubinstein, Dancy, & Korn, 2005). This design was selected because it has two assumptions that align with expectations of education interventions: 1) the benefit of intervention may plateau such that higher doses do not result in greater gains, and 2) harm does not increase with greater doses. In traditional drug trials, this harm is toxicity; in spelling instruction, such harm is not expected. Determining adequate dose will result in teacher guidance for dose spelling instruction that is maximally effective without using classroom teaching time that provides no added benefit and could be used for other instructional targets. This design has been used previously for educational intervention (word learning) dose studies (Storkel et al., 2017). Three concurrent treatment studies will address the specific aims. Analysis will determine which treatment intensity of 12, 24, 48, or 72 exposures is most effective for improving spelling for first and second grade students of varying spelling skill levels.

Participants

One-hundred twenty 1st and 2nd grade students will participate in the proposed project. Participants will speak English as their primary language and will exhibit average or above average nonverbal intelligence, measured by the *Test of Nonverbal Intelligence – 4th Edition* (Brown, Sherbenou, & Johnsen, 2010). Additionally, participants will have normal hearing per parent report. One-third of participants will have strong overall spelling skills, operationally defined as scoring above the average range (i.e., > 115 standard score) on the *Test of Written Spelling – 5th Edition* (Larsen, Hammill, & Moats, 2013); one-third of participants will have average overall spelling skills, operationally defined as scoring within the average range (i.e., 85 - 114 standard score) on the TWS-5; and one-third of participants will have poor overall spelling skills, operationally defined as scoring below the average range (i.e., < 85 standard score) on the TWS-5. Finally, participants will score below 50% accuracy on the pre-test of spelling of the long-vowel patterns to be targeted in the study.

Procedures

All study procedures will be completed using Zoom for Healthcare, a HIPAA-compliant video conferencing application. Schools will be provided with computers and headphones to be used in intervention and will keep these as a thank-you following study completion. Figure 1 displays the random assignment of participants to conditions and blocks within the escalation design. First, students will be categorized as strong, average or poor spellers, as described above. Next, three escalation trials will be run simultaneously – one for each spelling skill level group. The procedures described below apply to each of the three escalation trials.

Random assignment within blocks. Within each of the spelling skill level groups, students will be randomly assigned in blocks of four to one of the dose conditions (Figure 1). The dose conditions will be 12 exposures, 24 exposures, 48 exposures, and 72 exposures. Each session will contain 4 exposures per spelling pattern, and all conditions will consist of three sessions per week; the 12 exposure groups will last one week (3 sessions), 24 exposures will last two weeks (6 sessions), 48 exposures will last four weeks (12 sessions), and 72 exposures will last six weeks (18 sessions). Six spelling patterns of long vowel sounds will be targeted. Only one block will be active at a time, and a week in between blocks will be used for decision making and pretesting of the next block (if needed). We will recruit participants for up to 10 blocks, based on previous dose research in educational

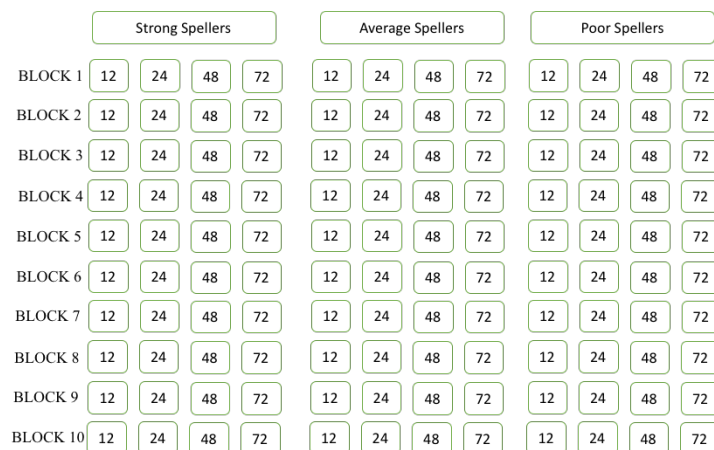


Figure 1. Dose Escalation Design. Each block contains 4 students from each group.

interventions (Storkel et al., 2017). Thus, a block of 12 participants (4 strong spellers, 4 average spellers, and 4 poor spellers) will participate in study activities at a time; one participant from each group will be assigned to the 12 exposures condition, one from each group will be assigned to the 24 exposures condition, one from each group will be assigned to the 48 exposures condition, and one from each group will be assigned to the 72 exposures condition. After the block is complete, the research team will meet to make decisions about dose escalation before commencing the next block.

Decision-making. At the conclusion of each block, the senior research team (Duncan, Werfel, Buell) will meet to complete the decision-making process. We will separately make decisions for each group. First, each participant within a group will be designated as a responder or non-responder to the instruction. Responders will be operationally defined as students who performed at 80% accuracy or higher for at least 4 of the targeted spelling patterns. Any student who does not meet this criterion will be classified as a non-responder. A pattern will be considered to have been established when it is replicated across at least three blocks. Therefore, after Blocks 1 and 2, the decision-making meeting will be a debriefing of findings from that block. Beginning at Block 3, we will examine the data for a consistent pattern of findings across three blocks within a single group. Once a consistent pattern has been identified, the escalation trial will be discontinued for that group. The figure below shows a study timeline of the project.

Weeks 1-4	Week 5	Weeks 6-11	Week 12	Weeks 13-18	Week 19	Weeks 20-25	Week 26	Weeks 27-32	Week 33	Weeks 34-39
Inclusionary and Descriptive Measures	Random Assignment of Students to Block 1 Pretest Block 1	Block 1 Intervention Posttest Block 1	Decision-Making Random Assignment of Students to Block 2 Pretest Block 2	Block 2 Intervention Posttest Block 2	Decision-Making Random Assignment of Students to Block 3 Pretest Block 3	Block 3 Intervention Posttest Block 3	Decision-Making Random Assignment of Students to Block 4 Pretest Block 4	Block 4 Intervention Posttest Block 4	Decision-Making Random Assignment of Students to Block 5 Pretest Block 5	Block 5 Intervention Posttest Block 5

Week 40	Weeks 41-46	Week 47	Weeks 48-53	Week 54	Weeks 55-60	Week 61	Weeks 62-67	Week 68	Weeks 69-74
Decision-Making Random Assignment of Students to Block 6 Pretest Block 6	Block 6 Intervention Posttest Block 6	Decision-Making Random Assignment of Students to Block 7 Pretest Block 7	Block 7 Intervention Posttest Block 7	Decision-Making Random Assignment of Students to Block 8 Pretest Block 8	Block 8 Intervention Posttest Block 8	Decision-Making Random Assignment of Students to Block 9 Pretest Block 9	Block 9 Intervention Posttest Block 9	Decision-Making Random Assignment of Students to Block 10 Pretest Block 10	Block 10 Intervention Posttest Block 10

Measures. Study 1 will include collection of inclusionary measures (described above in **Participants**), descriptive measures, and a curriculum-based outcome assessment. All assessors will be blinded to participant condition.

Descriptive measures. Parents will complete demographic questionnaires that include information such as student's gender, race, ethnicity, socioeconomic status, English-language learning status, and education history (e.g., repeated grades, presence of individualized education plans). These questionnaires will be completed using REDCap Survey. All students will complete the *Structured Photographic Expressive Language Test – 3rd Edition* (Dawson, Stout, & Eyer, 2003) to describe overall language abilities, the *Arizona Articulation Proficiency Scale-4th Edition* (Fudala & Stegall, 2017) to describe speech production abilities, and the *Woodcock Reading Mastery Test – III* (Woodcock, 2011) to describe overall reading abilities. These measures will be completed virtually, using Zoom for Healthcare. Dr. Werfel's lab has validated virtual assessment, with high test-retest reliability with in-person testing: language (.846), speech production (.955). Standard scores will be calculated according to published test manual instructions.

Curriculum-based outcome assessment. The curriculum-based outcome assessment will be a 60-item spelling test that measures spelling of words that contain the target vowel sound spellings used in the intervention condition, 10 words with each vowel spelling pattern targeted. The outcome assessment will be completed at three timepoints: pre-test, post-test, and maintenance. Scoring will be based on the vowel spelling rather than the whole word. Each word will receive a score of 1 for a correct vowel spelling or 0 for an incorrect vowel spelling. This assessment will be administered in groups for each block via Zoom for Healthcare, and schools will be provided stamped, addressed envelopes in which to return children's spellings.

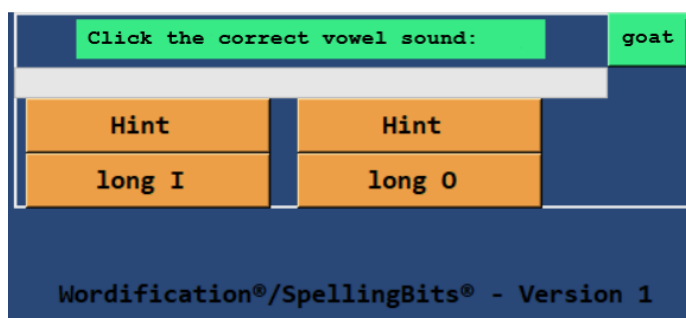
Intervention. *Wordification* provides instruction through games that utilize evidence-based principles of word study to teach students the principles of English spelling. The word study approach to spelling instruction is a systematic instructional approach that teaches children to notice the patterns of English spelling (Bear, Invernizzi, Templeton, & Johnston, 2011). It aims to help children develop general knowledge about the English spelling system (i.e., spelling patterns) and to increase specific word knowledge (i.e., spelling of individual words). *Wordification* will include games that follow principles of word sorting and word building. The beginning of each unit will include instruction of the target structure or rule and modeling of word sorts and word building activities that utilize that target. The application is programmed in Python, and the lexical database (accessed from Python) is programmed using MySQL. The database encodes each word's spelling, graphemic, phonetic, and syllabic properties, along with illustrative examples of use. Program assets include sound files that are used to populate the audio script of the game. The game itself runs on a university server and is accessed through a graphical user-interface (GUI). For this project, the GUI provides to student subjects, in each of the several schools participating in the study, anonymized user login to the server and access to the intervention. Each student subject's activity (time spent, responses to items, and error rates) is tracked and recorded, thus allowing the investigators the ability to monitor and assess each individual's progress and attainment. The webification of the application will be especially important in the coming school year, in light of our current Covid-19 circumstances, as it will allow the investigators to conduct the study remotely, without having to visit the participating classrooms. We anticipate introducing classroom teachers to the study protocols through online Zoom sessions and checking in with them periodically in the same fashion.

Word sorting. Word sorting is an analytic phonics approach to instruction that involves categorizing words by linguistic features. The act of sorting encourages students to search, compare, contrast, and analyze the linguistic features of words, and helps students to organize what they already know about specific words and apply that knowledge to new words. In each type of exercise, correct responses are rewarded, and incorrect responses result in a brief re-teaching episode and a prompt for the student to self-correct his/her response.

Children will be randomly assigned to one of the four dose conditions (12, 24, 48, 72). In the intervention, children will perform a sorting exercise involving two vowel sounds and six graphemic patterns (3 for each sound). They will first be asked to select the correct vowel sound for a target word, then asked to select the correct spelling pattern for that vowel sound, and finally asked to spell the target word. Subjects will sort 24 target words each into two vowel classes (long "I" and long "O"), with 12 examples of each, and will then sort the 12 examples for each vowel class into one of 3 spelling patterns ("igh", "y" and "i-consonant-e" for long "I"; "ow" "oa", and "o-consonant-e" for long "O"). For illustration, we can describe the intervention's presentation of the word "goat":

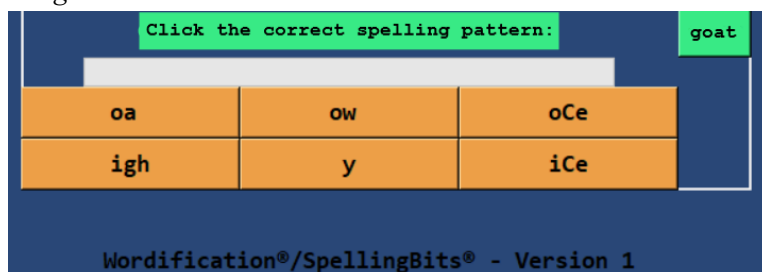
Game: "Let's figure out the vowel sound in *goat*. Here's a sentence with *goat* in it: *A goat is a farm animal. Goat* has three sounds: [g], [ow], and [t]. Which is the vowel sound in *goat*?"

[Student is presented with "Long I" and "Long O" buttons to choose from, along with a "Hint" button for each. If student clicks the correct button, the game says, "That's right. The vowel sound in *goat* is [ow]."
If student clicks the wrong button, the game says, "That's not right. The vowel sound in *goat* is not [ay]. Try again."
Student can commit two errors before the game moves them on.]



Game: "Now, pick the spelling pattern for the [ow] sound in *goat*."

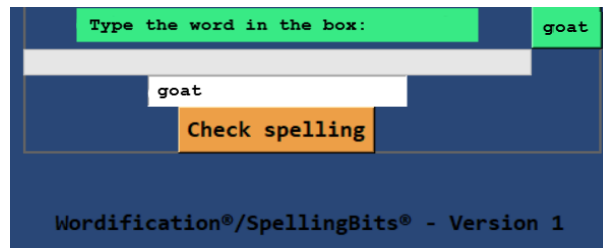
[Student is presented with a choice of six grapheme buttons: "oa", "ow", "o_e", "y", "igh", and "i_e". If subject clicks the correct button, the game says, "That's right. The [ow] sound in *goat* is spelled with o-a." If subject clicks the wrong button, the game might say, depending on the choice, "That's not right. The [ow] sound in *goat* is not spelled with o-



consonant-e. Try again,” or “That’s not right, i-g-h is not a spelling pattern for [ow]. Try again.” Student can commit two errors before the game moves them on.]

Game: “Now, type the word *goat* in the box.”

[Student types word. If word is spelled correctly, the game says, “That’s right, *goat* is spelled g-o-a-t. Good job.” If incorrect, the game says, “Sorry. Try again. Remember that *goat* has the long-o sound [ow] spelled with o-a.” Student can commit two errors before the game moves them on.]



Data Analytic Procedure

The analysis for Aim 1 is in line with a dose escalation model evaluating treatment intensity (used by Storkel et al., 2017; Hunsberger et al., 2005). To determine optimal treatment intensity for each skill, benchmarks for adequate progress must be set and evaluated. Post-test scores will be evaluated for each child, and responses will be classified as response or no response, as described in “Decision Making” above. Number of responders versus non-responders will be compared for each intensity. A follow-up repeated measures analysis of variance using pre-test score as a covariate and treatment intensity group membership as the between-subjects independent variable will evaluate, at the group level, overall differences in response by treatment intensity.

The analysis for Aim 2 will evaluate the predictive value of child characteristics on variation in treatment response across intensities. Child-level predictors will include linguistic knowledge at study onset (as indicated by descriptive measures) and demographic variables inclusive of maternal education and family history of reading disorders. To assess the effects of child-level predictors on child response to Wordification, multiple regressions will be run evaluating post-test spelling performance using a dummy-coded variable for intensity and continuous values of predictors (except family history will be dummy-coded). Initial correlations between predictors and outcome variables will be checked for all possible predictors, and those predictors significantly related to outcome variables will be entered into the regression analysis to assess moderation of treatment effect. The product term of the intensity and predictor for the dependent variables will be calculated to determine moderated treatment effect, consistent with the methodology described by Yoder & Compton (2004).

Power Analysis. Power was calculated using G*Power (Faul et al., 2009). A recent meta-analysis (Graham & Santangelo, 2014) indicated effect sizes for spelling instruction of $d = 0.54$ on spelling outcomes. With traditional assumptions ($\alpha = 0.05$, power = 0.80), a total sample size of 120 students will be powered to detect an effect size of $d = 0.26$.

Expected Outcomes

The working hypothesis is that children who receive higher intensities of *Wordification* instruction will exhibit greater gains in spelling skills than children who receive lower intensities, and that the optimal intensity for treatment gain will differ across groups. Specifically, we anticipate that strong spellers will be successful at 12 or 24 dose levels and additional intensity will not be necessary for gain. For average spellers, we anticipate that 24 or 48 dose levels will be optimal, and for poor spellers, we anticipate that 48 or 72 dose levels will be necessary for treatment gains. We additionally anticipate that child-level factors may moderate treatment outcomes when examining factors such as language and reading skills, as well as maternal education and family history of reading disorders. Alternately, it may be that the same dose level emerges as optimal across multiple or all groups. In any case, the findings will provide guidance for the further development of computer-based linguistic spelling instruction that is maximally effective across groups of spellers but does not needlessly use instructional time with no additional benefit to spelling skills that could be used to address other educational targets.

Potential Problems and Alternative Strategies

The primary potential problems in this study include attrition and adherence to study condition. To control for attrition, participants in all conditions will receive a \$25 gift card as a thank you for completing the study. To monitor problems with adherence to study conditions, study personnel will conduct monthly procedural fidelity observations and will debrief with teachers as needed. Additionally, teachers assigned to the control condition will receive *Wordification* free of charge at the conclusion of their participation in the study and be allowed to keep the study computers and headphones provided for the study for use in their classrooms. To allow for completion of the project in the uncertainty of the current COVID-19 pandemic, we have ensured that all study procedures can be completed virtually and via contactless drop-off and pick-up.